

Analysis of a candidate gene for the
control of floral heteromorphy in
Primula vulgaris

Olivia Victoria Constance Kent

Submitted for the Degree of Doctor of Philosophy

University of East Anglia

John Innes Centre

Department of Cell and Developmental Biology

September 2016

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract:

Primula vulgaris is a model species for the study of heterostyly, and displays two floral morphologies, pin and thrum, which show a reciprocal arrangement of the anthers and stigma. The differences in floral morphology are controlled by the *S* locus, which consists of several closely linked genes.

One of the first genes to be identified as part of the *S* locus was *GLO^T*, a paralogue of *GLOBOSA*. Preliminary results have shown that *GLO^T* expression is confined to the second and third floral whorls of thrum flowers. In pin flowers, where *GLO^T* is not expressed, the anthers are lowered. This study involves characterisation of the expression dynamics of *GLO^T* in comparison to its paralogue *GLO*, in the context of the recent discovery that the *S* locus is hemizygous and not heterozygous as previously thought.

The selection of normalisation genes for qPCR was conducted, and the temporal expression of both *GLO* and *GLO^T* was measured across bud development; the genes showed different expression patterns. RNA *in situ* hybridisation was then used to assess spatial expression of both genes in floral meristems, with *GLO* showing defined localisation within the developing second and third whorls and *GLO^T* showing more dispersed expression.

The interactions of *P. vulgaris* *GLO* and *GLO^T* proteins with *A. thaliana* MADS box proteins were tested in Yeast 2-Hybrid experiments, and while *GLO* showed interactions with the orthologue of its partner, AP3, *GLO^T* did not show interactions with any of the other proteins tested. Antibodies were designed against peptide sequences to assess protein localisation for use in future experiments.

This work has furthered knowledge on the expression patterns of these genes, the divergence of *GLO^T* from *GLO*, and has generated tools that will enable further analyses of the differences between these two genes.

Table of contents:

Abstract:	ii
List of figures:	x
List of tables:	xii
Acknowledgements:	xiii
CHAPTER 1: Introduction	1
1.1 Floral heteromorphy and heterostyly	1
1.1.1 Heterostyly and self-incompatibility	4
1.2 Heterostyly in <i>Primula</i> species	4
1.3 The <i>S</i> locus in <i>Primula</i> species.....	7
1.3.1 The <i>GPA</i> model for <i>S</i> locus structure	9
1.3.2 The use of floral mutants in the categorisation of the <i>S</i> locus.....	10
1.4 Floral organ identity and the ABC model	11
1.4.1 Identification of the A, B and C function genes	13
1.4.2 Revisions to the ABC model.....	14
1.5 Mutants of <i>P. vulgaris</i> and their links to the <i>S</i> locus	17
1.5.1 <i>Hose-in-Hose</i>	17
1.5.2 <i>sepaloid</i>	17
1.5.3 <i>Oakleaf</i>	18
1.6 Molecular markers for the <i>P. vulgaris S</i> locus	18
1.6.1 <i>SLP1</i>	18
1.6.2 <i>SLL1</i> and <i>SLL2</i>	19
1.6.3 <i>GLO</i>	20
1.7 Genetic map of the <i>S</i> locus	20
1.8 Genomic sequencing of wild type <i>P. vulgaris</i>	23

1.9 The genes within the <i>P. vulgaris</i> <i>S</i> locus.....	26
1.9.1 The duplication of <i>P. vulgaris</i> <i>GLO</i>	26
1.10 Experimental aims.....	27
 CHAPTER 2: Materials and methods	29
2.1 List of suppliers.....	29
2.2 Stock solutions and media.....	32
2.2.1 Chemicals	32
2.2.2 Stock solutions	32
2.2.3 Stock buffers	33
2.2.4 Growth media	34
2.3 Plant material	35
2.4 Seed sterilisation and germination.....	35
2.4.1 <i>Primula</i> species.....	35
2.4.2 Standard <i>A. thaliana</i> seed sterilisation.....	35
2.4.3 Sterilisation of transformed <i>A. thaliana</i> seeds	36
2.4.4 <i>A. thaliana</i> growth conditions.....	36
2.5 Nucleotide isolation	36
2.5.1 Genomic DNA extraction	36
2.5.1.1 illustra Nucleon PhytoPure Extraction Kit	37
2.5.1.2 QIAGEN DNeasy Plant Mini Kit	37
2.5.1.3 Edwards genomic DNA extraction	38
2.5.2 RNA isolation.....	39
2.5.3 Precipitation of RNA	40
2.5.4 cDNA synthesis.....	40
2.5.5 DNase treatment of RNA	40
2.5.6 cDNA synthesis for standard PCR	41

2.5.7 cDNA synthesis for quantitative real-time PCR.....	41
2.6 Polymerase Chain Reaction.....	42
2.6.1 Oligonucleotide design and synthesis.....	42
2.6.2 Standard PCR.....	42
2.6.3 Long and proofreading PCR.....	42
2.6.4 Gel Electrophoresis.....	43
2.6.6 Isolation of DNA from PCR products	43
2.6.7 QIAquick Gel Extraction Kit.....	44
2.6.8 QIAquick PCR Purification Kit.....	44
2.6.9 DNA sequencing	44
2.6.10 A-tailing of PCR products	45
2.7 Quantification of transcript by quantitative real-time PCR (qPCR)	45
2.7.1 The standard qPCR reaction.....	45
2.7.2 Testing of all primers for qPCR	46
2.7.3 Selection and testing of genes for normalisation of qPCR results	46
2.8 Molecular cloning.....	47
2.8.1 Ligation reaction for pGEM-T Easy constructs	47
2.8.2 Preparation of DH5 α competent cells.....	47
2.8.3 Transformation of DH5 α competent cells with pGEM-T Easy constructs	48
2.8.4 Ligation reaction for Gateway constructs	48
2.8.5 Transformation of TOP10 competent cells with Gateway constructs	48
2.8.6 Overnight liquid culture of cells	49
2.8.7 Plasmid purification	49
2.8.8 Gateway LR Clonase II reaction.....	50
2.8.9 Type IIS restriction enzyme cloning.....	51
2.8.9.1 Primer design for products to be used in Type IIS restriction enzyme cloning.....	51
2.8.9.2 Digestion and ligation reaction	53

2.8.9.3 Analysis of constructs	53
2.8.9.4 Transformation of <i>Agrobacterium tumefaciens</i>	53
2.8.9.5 Transformation of <i>A. tumefaciens</i> strain GV3101	54
2.8.9.6 Transformation of <i>A. tumefaciens</i> strain AGL1	54
2.9 <i>A. thaliana</i> transformation.....	54
2.9.1 Analysis of transformants	55
2.10 <i>In situ</i> hybridisation	55
2.10.1 Preparation of plasmids for probe synthesis.....	55
2.10.2 Linearisation of plasmids for probe synthesis	56
2.10.3 Probe-labelling reaction	56
2.10.4 Carbonate hydrolysis of probes.....	56
2.10.5 Dot blot testing probes	57
2.10.6 Preparation of tissue fixative.....	57
2.10.7 Fixation of plant tissue.....	58
2.10.8 Processing and embedding of fixed tissue	58
2.10.9 Embedding and sectioning of tissue	58
2.10.10 Pre-hybridisation.....	59
2.10.11 Hybridisation	59
2.10.12 Post-hybridisation	60
2.11 Sequence analysis	61
2.12 Software packages	61

CHAPTER 3: Using qPCR to monitor the quantitative changes in gene expression seen across the developmental stages of buds in <i>Primula vulgaris</i>	62
3.1 Introduction.....	62
3.2 <i>GLO^T</i> in <i>P. vulgaris</i>	64
3.3 <i>GLO^T</i> in <i>P. veris</i> and <i>P. elatior</i>	66

3.4 Selection of <i>P. vulgaris</i> flower bud sizes for quantitative real-time PCR (qPCR).....	69
3.5 The selection of appropriate reference genes for qPCR	69
3.5.1 Identifying the <i>P. vulgaris</i> homologues of <i>A. thaliana</i> normalisation genes	72
3.5.2 Testing the <i>P. vulgaris</i> homologues for specificity	73
3.6 The potential for normalisation	77
3.6.1 Analysis of raw Cq values	77
3.6.2 Δ Ct method of normalisation gene selection	80
3.6.3 NormFinder	82
3.6.4 Biogazelle qbase+	84
3.6.5 Conclusions that can be drawn from all methods.....	87
3.7 Testing the candidate normalisation genes.....	88
3.8 Assessment of inter-run variation	92
3.9 The expression patterns of <i>GLO^T</i>	95
3.10 Quantitative expression of <i>GLO</i>	96
3.11 Comparative expression of <i>GLO</i> and <i>GLO^T</i>	98
3.12 Summary of findings.....	98
 CHAPTER 4: Spatial localisation of <i>GLO</i> and <i>GLO^T</i> in <i>Primula vulgaris</i>	99
4.1 Introduction.....	99
4.1.1 Experimental aims	100
4.2 RNA <i>in situ</i> localisation.....	100
4.2.1 RNA <i>in situ</i> localisation of <i>GLO</i> in floral meristems.....	101
4.2.2 RNA <i>in situ</i> localisation of <i>GLO^T</i> in floral meristems	104
4.2.3 RNA <i>in situ</i> localisation of <i>GLO</i> and <i>GLO^T</i> in mature flower bud tissue	106
4.3 Immunolocalisation using peptide antibodies.....	108
4.3.1 Design of peptide antibodies for <i>GLO</i> and <i>GLO^T</i>	108
4.3.2 Design of peptide antibodies for other genes in the <i>S</i> locus	110

4.3.3 <i>In silico</i> testing of peptide specificity.....	112
4.4 Summary of findings.....	114
CHAPTER 5: Expressing <i>Primula vulgaris</i> MADS box genes in <i>Arabidopsis thaliana</i>	115
5.1 Introduction.....	115
5.1.1 Experimental aims	116
5.2 Promoter-reporter constructs design for <i>GLO</i> , <i>GLO^T</i> and <i>PI</i>	118
5.2.1 Construct design for promoter-reporter experiments.....	118
5.2.2 Promoter-reporter construct production	121
5.2.3 Analysis of transformants for promoter-reporter constructs	121
5.3 <i>In silico</i> promoter analysis of <i>GLO</i> and <i>GLO^T</i>	123
5.4 The interactions of <i>P. vulgaris</i> and <i>A. thaliana</i> MADS box transcription factors	129
5.4.1 The interaction of <i>P. vulgaris</i> and <i>A. thaliana</i> genes in yeast	129
5.5 Complementation of the <i>A. thaliana</i> <i>pi-1</i> mutant with <i>P. vulgaris</i> genes	133
5.5.1 Construct design for complementation of the <i>pi-1</i> mutant	134
5.5.2 Construct production for the complementation of the <i>pi-1</i> mutant.....	134
5.5.3 Analysis of <i>pi-1</i> complementation transformants.....	138
5.6 Overexpression of <i>GLO</i> in <i>A. thaliana</i>	138
5.6.1 Construct production for the overexpression of <i>GLO</i>	138
5.6.2 Analysis of <i>GLO</i> overexpression transformants	141
5.7 Summary of findings.....	141
CHAPTER 6: Discussion	142
6.1 Introduction.....	142
6.1.1 The duplication of <i>GLO</i>	143
6.2 The selection of genes for normalisation of <i>GLO</i> and <i>GLO^T</i> expression.....	143
6.3 Temporal expression of <i>GLO</i> and <i>GLO^T</i> expression	144

6.4 Localisation of <i>GLO</i> and <i>GLO^T</i> : RNA <i>in situ</i> hybridisation	146
6.5 Localisation of <i>S</i> locus proteins: design of peptide antibodies.....	147
6.6 Interactions of GLO and GLO ^T with other proteins	148
6.7 Promoter analysis of <i>GLO</i> and <i>GLO^T</i>	149
6.8 Design of constructs for transformation of <i>A. thaliana</i> with <i>P. vulgaris</i> genes	149
6.8.1 Promoter-reporter constructs.....	149
6.8.2 Complementation of the <i>pi-1</i> mutant with <i>P. vulgaris</i> genes	150
6.8.3 Overexpression of <i>GLO</i> in <i>A. thaliana</i>	150
6.9 Conclusions.....	152
APPENDIX.....	153
REFERENCES.....	170

List of figures:

Figure 1.1: A generic example of heterostyly.....	3
Figure 1.2: Heterostyly in <i>P. vulgaris</i>	5
Figure 1.3: A comparison of <i>P. vulgaris</i> floral phenotypes	8
Figure 1.4: The ABC model of floral development	12
Figure 1.5: The updated ABC model of floral development	16
Figure 1.6: The <i>P. vulgaris</i> <i>S</i> locus in relation to known markers	22
Figure 1.7: Comparison of the <i>S</i> and <i>s</i> haplotypes of the <i>S</i> locus	25
Figure 2.1: The recognition site and overhangs that result from digestion with BsaI	52
Figure 2.2: Removal of an unwanted BsaI site	52
Figure 3.1: Expression of <i>GLO</i> and <i>GLO^T</i> in the flower buds of <i>P. vulgaris</i>	65
Figure 3.2: Comparison of the amino acid sequences of <i>GLO</i> and <i>GLO^T</i> in <i>P. vulgaris</i>	65
Figure 3.3: Expression of <i>GLO</i> and <i>GLO^T</i> in <i>Primula</i> flower buds	67
Figure 3.4: Alignment of the amino acid sequence of <i>GLO</i> in three <i>Primula</i> species	68
Figure 3.5: Alignment of the amino acid sequence of <i>GLO^T</i> in three <i>Primula</i> species	68
Figure 3.6: Example primer efficiency check and melt peak analysis	75
Figure 3.7: Comparison of geNorm M values	85
Figure 3.8: <i>GLO^T</i> expression normalised against <i>PP2A</i> and <i>TUA</i>	89
Figure 3.9: <i>GLO^T</i> expression normalised against <i>ELF1α</i> , <i>PP2A</i> and <i>TUA</i>	89
Figure 3.10: Expression of <i>GLO^T</i> across the developmental stages of <i>P. vulgaris</i>	91
Figure 3.11: Expression of <i>GLO^T</i> across the developmental stages of <i>P. vulgaris</i>	91
Figure 3.12: Expression of <i>GLO^T</i> across the developmental stages of <i>P. vulgaris</i>	93
Figure 3.13: Expression of <i>GLO^T</i> across the developmental stages of <i>P. vulgaris</i>	93
Figure 3.14: Expression of <i>GLO</i> across the developmental stages of <i>P. vulgaris</i>	97
Figure 3.15: Expression of <i>GLO^T</i> in comparison to <i>GLO</i>	97
Figure 4.1: ClustalO alignment of <i>GLO</i> and <i>GLO^T</i> cDNA sequences	102
Figure 4.2: RNA in situ localisation of <i>GLO</i> in <i>P. vulgaris</i> floral meristems	103
Figure 4.3: RNA <i>in situ</i> hybridisation of <i>GLO^T</i> probes on thrum floral meristems	105

Figure 4.4: RNA <i>in situ</i> hybridisation of <i>GLO^T</i> probes on pin floral meristems	105
Figure 4.5: Thrum 5 mm flower bud sections	107
Figure 4.6: Position of the peptides within the protein sequence for GLO and GLO ^T	109
Figure 4.7: Position of the peptide within the protein sequence of KFB ^T	111
Figure 4.8: Position of the peptides within the protein sequence of PUM ^T	111
Figure 5.1: The four promoter-reporter constructs	120
Figure 5.2: Example of the analyses performed on the promoter-reporter constructs	122
Figure 5.3: Alignment of the <i>GLO</i> and <i>GLO^T</i> promoter sequences.....	125
Figure 5.4: MADS box trascription factor binding sites within <i>GLO</i> promoter	127
Figure 5.5: MADS box trascription factor binding sites within <i>GLO^T</i> promoter	128
Figure 5.6: Yeast 2-Hybrid data for the AP3 and PI AD- and BD-fusions	132
Figure 5.7: The six complementation constructs	136
Figure 5.8: Example of the analyses performed on the complementation constructs	137
Figure 5.9: Example of the analyses performed on the GLO overexpression construct	140

List of tables:

Table 3.1: Initial list of candidates for gene normalisation	71
Table 3.2: Refined list of candidates for gene normalisation	74
Table 3.3: R^2 values for primer pairs of potential reference genes	76
Table 3.4: Raw Cq analysis of normalisation genes	79
Table 3.5: ΔCq analysis of normalisation genes	81
Table 3.6: NormFinder analysis of normalisation genes	83
Table 3.7: geNorm analysis of normalisation genes	86
Table 3.8: Standard deviation of qPCR performed on separate plates combined plates	94
Table 4.1: Peptide sequences used for antibody production	113
Table 5.1: Description of the MADS box protein binding sites within <i>GLO</i> promoter	127
Table 5.2: Description of the MADS box protein binding sites within <i>GLO^T</i> promoter	128

Acknowledgements:

I would first like to thank my supervisor Professor Phil Gilmartin for his support and guidance over the last four years, and without whom none of this would have been possible. Thanks also go to the other members of my supervisory panel: Professor Lars Østergaard and Professor Cathie Martin for their encouragement and helpful suggestions regarding the project, and to Dr. Jinhong Li for all of the practical help he has provided over the years.

On that note I would also like to thank Ali Pendle who has taught me so much about so many things, and who has always been willing to spend her tea breaks troubleshooting my technical problems. Her patience, generosity and help over the course of my PhD have been invaluable. I would also like to thank Pauline Stephenson for her technical help and support, particularly with the RNA *in situ* hybridisation, and Dr. Barry Causier for carrying out the Yeast 2-Hybrid experiments.

I would like to thank the other members of the Gilmartin lab, past and present, for their help and encouragement. Particular thanks go to Matt Smith for all of his help and advice in the first year of my PhD, Lizzy Ingle for her patience when answering all of my silly questions, and Jon Cocker for help with the bioinformatics.

I must also offer thanks to my friends, who have been so supportive and have shown me endless patience during the thesis-writing period. Thank you to Andy Iskauskas for always keeping me motivated and for proofreading this thesis, and to Anna Mudge, Martin Dixon, Jim Rowe and Dave Mentlak for always managing to make me laugh.

A special thank you goes to Jack Lee for his unwavering support, and for all of the time he has spent proofreading my work, listening to presentations, discussing experiments and making cups of tea. I couldn't have done it without you.

CHAPTER 1

Introduction

1.1 Floral heteromorphy and heterostyly

In the majority of hermaphroditic plant families individuals of the same species display one distinct floral morphology, subject to minor environmental variations. There are, however, a number of cases in which plants of the same species can show markedly different discrete hermaphroditic floral morphologies; this is known as floral heteromorphy. When the difference in floral morphology is specifically that the style length is different in relation to the anther position, this is referred to as heterostyly.

A plant is distylous if there are two flower forms, or tristylous if there are three; the positions of the style in relation to the anthers for both distyly and tristyly is shown in figure 1.1. Distylous species that have been studied previously include several members of the *Linum* genus (such as *Linum grandiflorum* and *Linum perenne*) (Darwin 1877), the *Turnera* genus (such as *Turnera ulmifolia*) (Barrett 1978) and numerous members of the *Primulaceae* family. Tristylous species include Purple Loosestrife (*Lythrum salicaria*) (Heuch 1980) and the Woody Liana (*Hugonia serrata*) (Thompson *et al.* 1996).

The arrangement of sexual organs in heterostylous species is such that male and female organs are spatially separated within the flower to promote out-breeding. The style and anthers are in reciprocal positions in different flower morphologies (figure 1.1), and this arrangement of sexual organs is known as reciprocal herkogamy (Webb and Lloyd 1986). The majority of heterostylous species rely on insect-mediated pollination, and a reciprocal arrangement maximises cross-pollination. Darwin (1877) suggested that

when an insect visits flowers of different morphs, pollen accumulates on different parts of the body, and it is this spatial separation of pollen that increases cross-pollination. Continuation of this work suggests that this is the case (Kohn and Barrett 1992, Lloyd and Webb 1992).

In addition to this increase in cross-pollination caused by the arrangement of the sexual organs, a number of heterostylous species also display a heteromorphic self-incompatibility system, which prevents or dramatically reduces the number of successful self- and intra-morph pollination events occurring. There are, however, exceptions such as the *Lithodora* and *Glandora* genera, in which the link between heterostyly and the self-incompatibility system has been found to be weak (Ferrero *et al.* 2011).

This combination of self-incompatibility mechanism and heterostyly was noticed by Darwin (1877) who, when performing crosses in *Primula* species, termed inter-morph pollinations ‘legitimate unions’, as these produced viable seeds, and crosses within an individual flower, or between flowers of the same morph ‘illegitimate unions’, as these produced either no seed, or less seed.

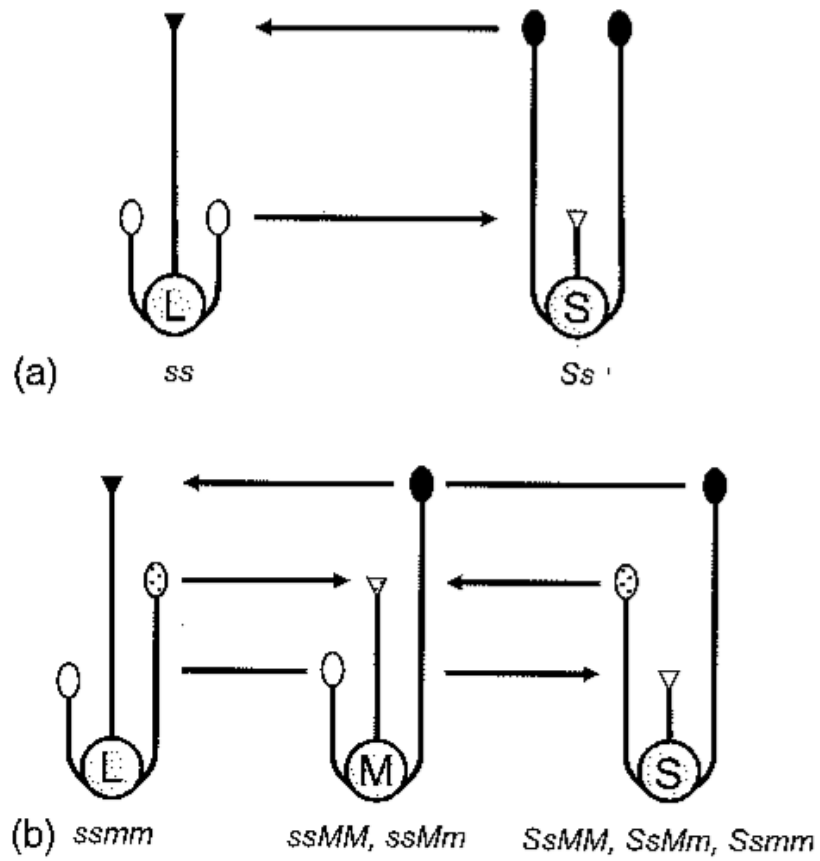


Figure 1.1: A generic example of heterostyly. Positions of the style and anthers differ between flower morphologies in distyly (a) and tristyly (b). Image from ‘Self-incompatibility in flowering plants’ Franklin-Tong, 2008.

1.1.1 Heterostyly and self-incompatibility

The region of the genome that is responsible for floral heteromorphy is referred to as the Style length locus (*S* locus). In self-incompatibility systems such as found in *Nicotiana*, *Brassica* and *Papaver*, the *S* refers to the Self-incompatibility locus. Self-incompatibility falls into one of two forms: gametophytic or sporophytic. In gametophytic self-incompatibility the haploid genome of the pollen determines its incompatibility phenotype, whereas in sporophytic self-incompatibility, the diploid genome of the parent plant determines the incompatibility phenotype of the pollen (Franklin-Tong 2008). The *Primulaceae* (discussed in section 1.2), *Brassicaceae* and *Lythraceae* are examples of families which exhibit sporophytic self-incompatibility, whereas the *Solanaceae*, *Rosaceae* and *Papaveraceae* are examples of families that utilise gametophytic self-incompatibility. (Franklin-Tong 2008; Poulter *et al.* 2010).

1.2 Heterostyly in *Primula* species

The *Primula* genus consists of over 90% heterostylous species, the highest percentage of any genus, and has been suggested for use as the model in the study of heterostyly (Mast and Conti 2006). As a distylous species *P. vulgaris* displays two floral morphologies, known as pin and thrum (Darwin 1877); these are shown using the *P. vulgaris* wild type in figure 1.2. The distinct differences between the two flower forms were first noticed by Clusius in 1583 (van Dijk 1943, Ganders 1979), however it was Darwin who made the majority of detailed observations as to the differences between the two forms, and the *Primula* species *Primula veris* formed a large proportion of his book ‘The Different Forms of Flowers on Plants of the Same Species’ (1877).

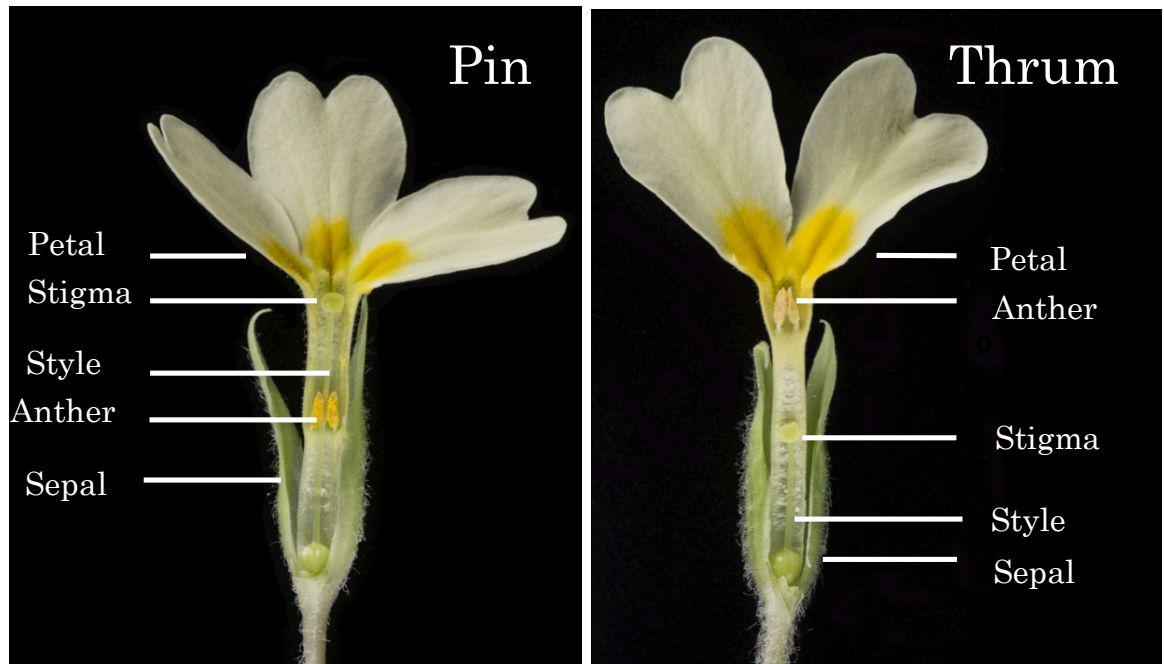


Figure 1.2: Heterostyly in *P. vulgaris*. Cross sections of mature pin and thrum *P. vulgaris* wild type flower buds show the difference in the position of the floral organs (labelled) between the flower morphs.

The most obvious physical difference between the two forms is the position of the style in relation to the anthers; in the pin form the stigma is seen at the mouth of the corolla tube and the anthers are situated half way up the corolla tube. In the thrum the anthers are seen at the corolla mouth and the style reaches to approximately half way up the corolla tube. Other physical differences that have been noted between the two forms include the different pollen sizes: each pin pollen grain is only $\frac{2}{3}$ of the size of a thrum grain, however the anthers are the same size in both floral morphologies (Darwin 1877). The styles of both morphologies do not only differ in length, but their stigmatic surfaces. The pin stigma displays a depressed centre and long stigmatic papillae cells, whereas the thrum stigma appears more rounded with shorter papillae cells (Darwin 1877). In his book, Darwin (1877) also analysed a number of *P. vulgaris* individuals, and made similar observations regarding the differences seen between pins and thrums.

The reciprocal herkogamy that is seen in *P. veris* and *P. vulgaris*, is as in the vast majority of cases, strongly linked to a heteromorphic self-incompatibility system. While self-pollination is possible in *Primula* species, the seed set is greatly reduced. When performing crosses on *P. veris* it was found that the pin self- or intramorph-pollination events yielded more seeds than thrum self- or intramorph pollinations (Darwin 1877). Though he was able to make a number of observations on the subject, Darwin was unable to provide an explanation for what was causing these physical differences, and differences in pollination efficiency between the morphologies, and it was only with the application of genetics that this became possible.

1.3 The *S* locus in *Primula* species

Following the observational work conducted predominantly by Darwin (1877), it was Bateson and Gregory (1905) who applied genetics to the heterostyly and self-incompatibility seen in *Primula*. Their work involved the cross-pollination experiments between *P. sinensis* and *P. acaulis* (the archaic name for *P. vulgaris*), and showed that the inheritance of heterostyly was Mendelian. They demonstrated that the thrum form contains one dominant and one recessive allele of the *S* locus, *Ss*, while the pin form is homozygous recessive, *ss*.

Between 1933 and 1936, Ernst performed crosses between *P. viscosa* and *P. hortensis*, and in doing so isolated progeny in which he believed changes had occurred within the *S* locus (Ernst 1933, 1936a 1936b). In his investigation, he came across what appeared to be the breakdown of heterostyly; homostyle individuals in which the anthers and stigma were at the same height. Short homostyles display a phenotype in which the anthers are in the position of the pin, and the stigma in the position of the thrum. Long homostyles display a phenotype where the anthers are in the position of the thrum and the stigma in the position of the pin (Ernst 1933, 1936a); images of the long and short homostyles in comparison to the pin and thrum flower morphologies are shown in figure 1.3. Not only are both sexual organs present at the same height, self-pollination of these individuals was considered a 'legitimate union' by Darwin's standards, as the anthers are of that of a thrum, and the style of that of a pin (Richards 2003).

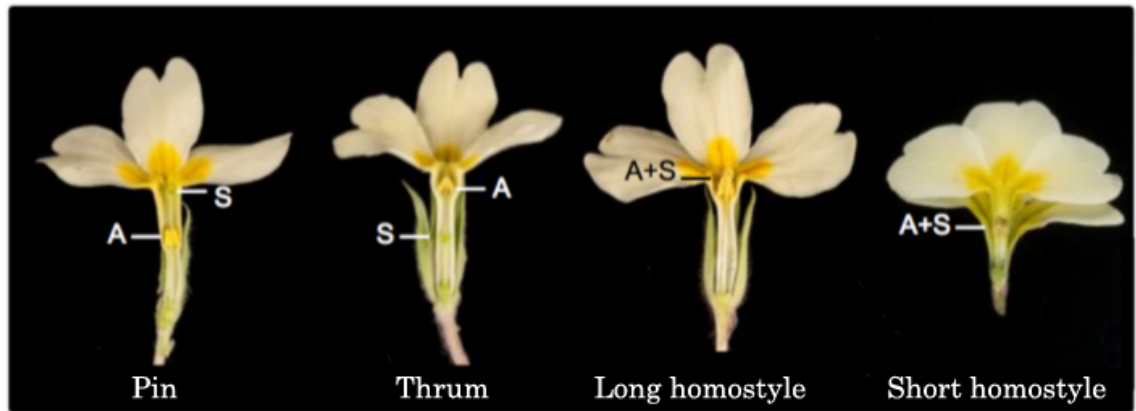


Figure 1.3: A comparison of *P. vulgaris* floral phenotypes. The height of the anthers (A) and stigma (S) are marked for each flower morph. The short homostyle flower also carries the *Hose-in-Hose* mutation (described in section 1.5.1), resulting in the conversion of sepals to petals. Image modified from Li *et al.* 2016.

Ernst believed that these phenotypes were caused by mutations occurring within the region of the genome controlling heterostyly, and defined three tightly-linked parts of the locus controlling heterostyly as *G*, *P* and *A* (Ernst 1936b, 1955). *G* is responsible for the difference in the height of the styles between the two, *P* is responsible for the different sizes of pollen, and *A* is responsible for the difference in anther height seen between the floral morphs (Ernst 1936b, 1955). The functions were later redefined as being somewhat more complex, with the *G* function also being proposed to be involved in the female aspect of the self-incompatibility system, and *A* function in the male aspects of self-incompatibility (Dowrick 1956; Richards 2003). This work brought about the beginning of the ‘supergene’ theory as to the inheritance of heterostyly, in which a number of tightly-linked genes work together to bring about certain characteristics (Lewis and Jones 1992).

The long homostyles phenotype described in Ernst’s work was found to exist in natural populations of *P. vulgaris*; Crosby discovered two populations of long homostyles in the wild, one in the Chilterns and one in Somerset (1940, 1949). Investigations into the floral morphologies found within these populations found that pins were found with a frequency of around 20%, whereas the remaining 80% were long homostyles, and thrums had been out-competed (Crosby, 1940, 1949). While Crosby believed that the ability of the long homostyles to self-fertilise was advantageous over the obligate outcrossing thrum (1958), an alternate theory proposed was that increased intramorph pollination, rather than self-pollination, was responsible for the success of the long homostyles within the population (Bodmer 1958); further research found the second theory to be correct (Bodmer 1960). Understanding the method by which the long homostyles spread throughout the population did not, however, explain the rarity of such wild populations, and Bodmer remarked that the selection for the long homostyles must be low (Bodmer 1960).

1.3.1 The *GPA* model for *S* locus structure

While it was Ernst (1936b, 1955) who suggested that there were three distinct elements required for heteromorphy and self-incompatibility to be present, *G*, *P* and *A*, it was Dowrick (1956) who expanded upon this idea and attempted to define the methods by which the long and short homostyles came into existence, subsequently using this

information to determine the order of the genes within the locus. Dowrick (1956) suggested that the order of these genes was crucial to their function, and that recombination events between these genes led to the different floral morphologies and levels of self-incompatibility observed by Ernst in 1936, not mutations as he had previously suggested.

The order of the genes present in the *S* locus, suggested by Dowrick (1956), was *GPA*; thrums were proposed to be heterozygotes, displaying the genotype *GPA/gpa*, and pins were proposed to be homozygous recessive, *gpa/gpa*. The dominant *G* gene found only in thrums was proposed to be responsible for the short style possessed by this floral morph and thus must be repressing style length, the dominant *P* for the large pollen, and the *A* for the high anther position within the corolla tube. This conclusion regarding the gene order was reached through the analysis of presumed rare recombination events occurring within the *S* locus (Dowrick 1956).

The genotype of long homostyle individuals was proposed to be *gPA/gpa*, and was assumed to have come about by a recombination event within the *S* locus. The genotype of short homostyles is proposed to be *Gpa/gpa* (Dowrick 1956). As previously noted both the long and short homostyles are self-compatible, and it was theorised that the breakdown of the *S* locus is responsible for this, but not been confirmed. As a result, efforts were made to locate the *S* locus within the *P. vulgaris* genome.

1.3.2 The use of floral mutants in the categorisation of the *S* locus

What is known about the *GPA* model has been gleaned from crosses, but this brought us no closer to the physical location of the *S* locus within the *P. vulgaris* genome. Three *P. vulgaris* mutants, one of which displays a floral and leaf phenotype, and two of which are homeotic mutants, appear to be linked to the *S* locus by way of their inheritance. As such, these mutants were proposed to be of importance in terms of finding the physical location of the *S* locus. The phenotype of the homeotic mutants is due to the ectopic expression or loss-of-function of one of the key genes involved in the development and maintenance of floral architecture; the mode for which is known as the ABC model (Coen and Meyerowitz 1991).

1.4 Floral organ identity and the ABC model

The majority of what is known about the development of floral organs is due to studies conducted in *Arabidopsis thaliana* and *Antirrhinum majus*. Work analysing floral homeotic mutants with prominent phenotypic differences from wild type led to the development of the original ABC model of flower development (Coen and Meyerowitz 1991). The *A. thaliana* floral mutants *apetala1* (*ap1-1*) and *apetala2* (*ap2-1*) display altered first and second whorls, *pistillata* (*pi-1*) and *apetala3* (*ap3-1*) both show the conversion of petals to sepals and stamen to carpels, and *agamous* (*ag-1*) shows the conversion of the stamen and carpels to sepals and petals. While no direct *A. majus* equivalent exists for the *ap-1* and *ap-2* *A. thaliana* mutants, the *globosa* (*glo*) and *deficiens* (*def*) mutants mirror the phenotypes seen in *pi-1* and *ap3-1*. Likewise, the *ag-1* mutant phenotype is mirrored by that of the *plena* (*plen*) mutant in *A. majus*. The simplest form of the ABC model, and how the above *A. thaliana* floral mutants relate to it, are shown in figure 1.4, where it is suggested that each floral whorl is controlled by the action of different genes, given the functional designations of A, B and C.

A function genes control the development of the sepal structures in the first floral whorl, the combination of A and B function genes give rise to petals in the second whorl, B and C function genes bring about stamen development in the third whorl, and the lone action of C function genes controls carpel development in the fourth whorl. A and C function genes are functionally antagonistic; in the absence of an A function gene a C function gene will take its place, and vice versa.

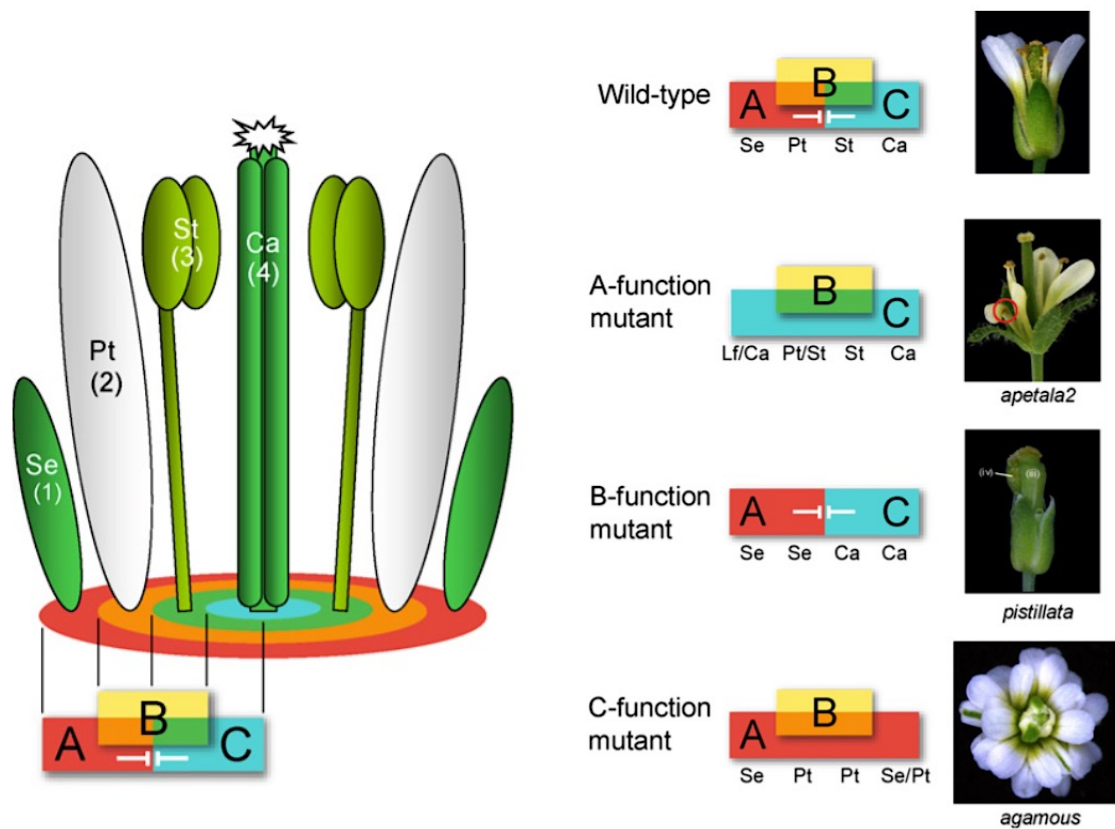


Figure 1.4: The ABC model of floral development. The whorls of the *A. thaliana* flower are marked: Se (1) = sepals, Pt (2) = petals, St (3) = stamens, Ca (4) = carpels. At the base of the floral diagram the colours represent the genes acting on this particular whorl, as shown in the block diagram below. To the right, the floral mutants seen when misexpression of the A B and C function occurs. Image adapted from Causier *et al.* 2010.

1.4.1 Identification of the A, B and C function genes

When the genes containing the mutations were identified, two A function genes were identified in *Arabidopsis*; *APETALA1* (*AP1*) (Mandel *et al.* 1992) which is a MADS box protein, and *APETALA2* (*AP2*) (Jofuku *et al.* 1994). Due to the lack of an *A. majus* homeotic mutant it was harder to identify the A function genes in this species, however *SQUAMOSA* (*SQUA*) shows the highest homology to *AP1* in *A. majus*. The genes containing the *ap3-1* and *pi-1* mutations were discovered; both *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) are MADS box transcription factors and perform the B function in the ABC model, as are *GLOBOSA* (*GLO*) and *DEFICIENS* (*DEF*) in *A. majus*. The gene containing the *ag-1* mutation, *AGAMOUS* (*AG*), appears to be the only C function gene in *A. thaliana*, and is also a MADS box transcription factor, as is its *A. majus* orthologue, *PLENA* (*PLE*).

From this, it is obvious that the majority of the genes involved in floral organ identity encode MADS box transcription factors. The MADS box family is so-called due to the similarity at the protein level of the DNA-binding domains of the *A. majus* genes *DEF* and *AG* to the *MINICHROMOSOME MAINTENANCE 1* (*MCM1*) gene in yeast, and *SERUM RESPONSE FACTOR* (*SRF*) in humans (Schwarz-Sommer *et al.* 1990); the acronym, MADS, utilises the first letter of each of these gene names. Two different types of MADS box protein have been identified; Type I, which show higher levels of similarity to the human *SRF*, and Type II, which show higher similarity to the more-recently discovered human gene *MYOCYTE ENHANCER FACTOR 2* (*MEF2*). The further identification of MADS box proteins in *A. majus* and *A. thaliana* have indicated that the MADS box genes of key importance in floral development can be categorised as Type II; for a review of the function of MADS box proteins and their function in floral development see Becker and Theißen 2003.

1.4.2 Revisions to the ABC model

More recent work has resulted in this model being re-evaluated, beginning with the discovery of a number of other genes that have an impact on floral organ identity. It has long been noted that there were aspects of the ABC model that did not fit; ectopically expressing B function and C function genes does not bring about the formation of floral organs outside of the flower (Krizek and Meyerowitz 1996). As a result, it was proposed that another gene, or group of genes, was involved in defining floral organ identity.

The *SEPALLATA* genes (*SEP1-4*) act in a semi-redundant fashion; a triple mutant of *sep1 sep2* and *sep3* results in the conversion of all floral organs to sepals; a quadruple mutant including *sep4* results in the conversion of all floral structures into leaf-like tissues (Pelax *et al.* 2000; Ditta *et al.* 2004). This would suggest that the *SEP* genes are required for the activity of the A, B and C functions genes in *A. thaliana*. This, combined with the discovery that a number of angiosperm species possess no confirmed A function genes (Litt and Kramer 2010) have resulted in a new model being proposed, with the A function being redefined and the E function *SEP* genes being included. They were named the E function as the D function had already been defined as the genes responsible for specifying ovule identity, first identified in *Petunia hybrida* (Colombo *et al.* 1995; Angenent and Colombo 1996) and subsequently in *A. thaliana* (Savidge *et al.* 1995; Flanagan *et al.* 1996; Pinyopich *et al.* 2003).

The new model, in which the E function is taken into account, is designated the ABCE, sometimes also referred to as the (A)BCE model (Davis *et al.* 2006; Causier *et al.* 2010). In this model the functions of the original A function and the more recently defined E function are combined as (A), in which this new (A) function specifies the identity of floral meristems not floral organs, in that it brings about the correct organisation of the floral organs by ensuring that the B and C function genes are expressed in the correct regions. In this model it is assumed that sepals are the default state for all tissue defined as being part of the floral meristem. The *ap1* and *ap2* mutants lack sepals, which would support the idea that they are responsible for defining tissue as floral meristem.

The revisions to the model have also altered the thinking about how MADS box proteins interact. It is known that the MADS box proteins involved in floral development form dimers to bring about their functions; immunoprecipitation experiments conducted in *A. thaliana*, have shown that AP1 and AG form homodimers, whereas PI and AP3 form

heterodimers (Reichman 1996). However, it has been suggested that the SEP proteins, specifically SEP3, is involved in forming higher-order complexes, some of which include the floral development MADS box proteins. When tested in yeast, the full-length proteins of PI and AP3 do not interact (Yang *et al.* 2003), however when the interactions of PI, AP3 and SEP3 are tested in yeast, interactions are seen, suggesting that SEP3 has a role in mediating the interactions between AP3 and PI (Immink *et al.* 2009). Experiments conducted *in planta* confirmed that SEP3 is able to affect the localization of the PI/AP3 complex, which supports the idea that these MADS box proteins are not just working as dimers. The quartet model suggests that the four different whorls of the flower are defined by the formation of heterotetramers (Theißen 2001), with experimental evidence available to suggest that PI and AP3 are incorporated into tetramers with SEP3 (Melzer and Theißen 2009). The revised ABCE/(A)BCE model is displayed visually in figure 1.5, including the proposed formation of the heterotetramers defining floral whorl identity.

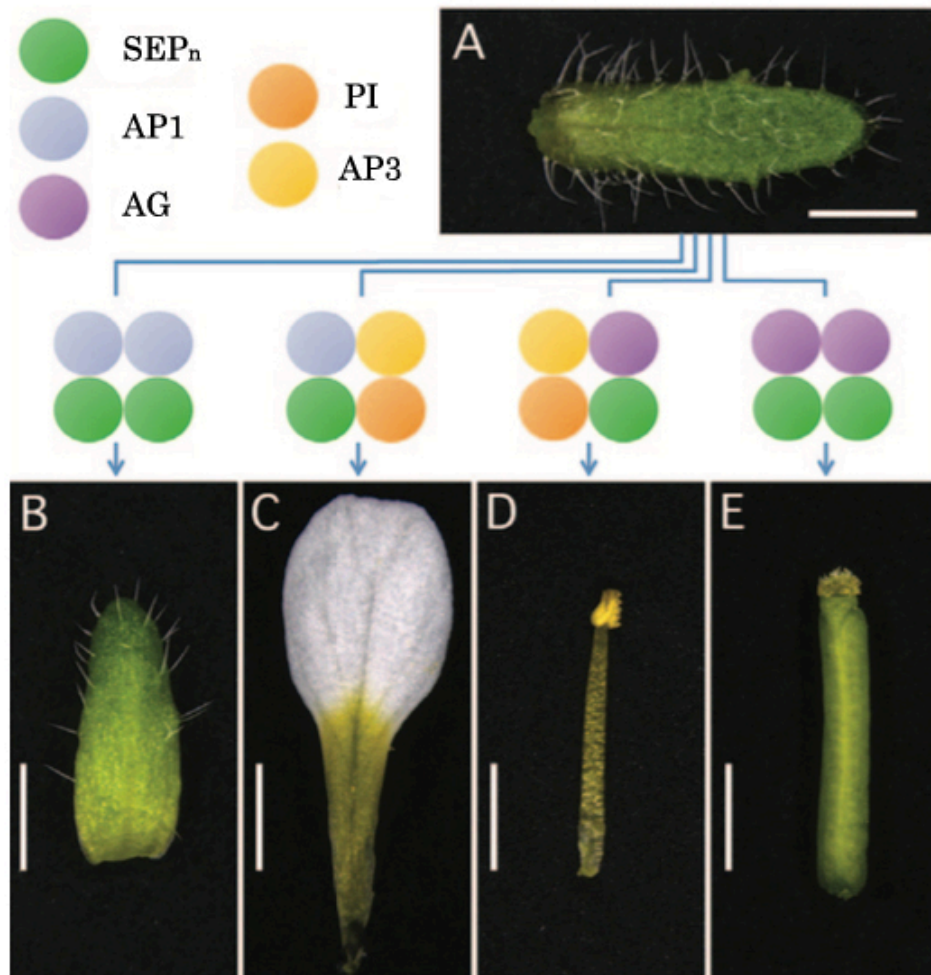


Figure 1.5: The updated ABC model of floral development. Coloured circles represent MADS-box proteins responsible for organ identity in *A. thaliana*. SEP_n represents multiple, partially redundant SEP proteins. Panel A represents the ground state of floral organs, represented by a cauline leaf. Panel B shows a sepal, C a petal, D a stamen and E a carpel. Scale bar: 1 mm. Image modified from Sablowski 2015.

1.5 Mutants of *P. vulgaris* and their links to the *S* locus

1.5.1 *Hose-in-Hose*

The *Hose-in-Hose* mutant was first acknowledged in the 16th Century (Gerard 1597), and is a dominant mutation (Webster and Grant 1990) in which ectopic expression of a B function gene results in the conversion of sepals to petals, however the level of conversion varies between individuals (Li *et al.* 2010). It was previously thought that the *Hose-in-Hose* mutation was completely linked to the thrum allele of the *S* locus, as all of the plants displaying this mutation were phenotypically thrum, however more recent research has found that it is possible for pin plants to display the *Hose-in-Hose* mutation (Webster and Grant 1990, Webster and Gilmartin 2003). The presence of the mutation in pins has been proposed to be due to a rare recombination event, occurring between *Hose-in-Hose* and the *S* locus, with the two loci displaying a linkage distance of 0.7 cM. The gene containing a mutation that results in this phenotype is the B function MADS box gene *GLOBOSA* (*GLO*) (Li *et al.* 2010), the orthologue of *A. majus GLO* and *A. thaliana PI* (Tröbner *et al.* 1992; Goto and Meyerowitz, 1994); the importance of this gene is discussed in sections 1.6.3 and 1.9.1.

1.5.2 *sepaloid*

The *sepaloid* mutation was discovered in 1996 (Webster and Gilmartin 2003), and presents as having a loss-of-function in a B function gene. The mutation results in the expression of an A function gene in regions of the flower that the B function would usually prevent, causing the conversion of petals to sepals in the weakest form of the mutation, and complete conversion of all four whorls at its most severe (Webster and Gilmartin 2003; Li *et al.* 2008). The gene containing a mutation that results in the *sepaloid* phenotype has not yet been identified. Analysis of fertile *sepaloid* mutants suggests it is linked to the pin allele of the *S* locus, with the two loci displaying a linkage distance of 0.3 cM (Webster and Gilmartin 2003, Webster 2005).

1.5.3 *Oakleaf*

The *Oakleaf* mutant was initially identified in 1999, and affects both the floral and leaf phenotypes (Webster 2005). The name comes from the lobed leaves that are associated with this mutant, which show similarities to those of *Quercus* species; the petals of plants carrying this mutation are also exaggeratedly lobed. The mutational severity varies, and the mutation is dominant, co-segregating with the pin allele of the *S* locus. The mapping distance between these two loci has been found to be 3.3 cM (Cocker *et al.* 2015). It is not known which gene is responsible for the appearance of these plants; recent work has suggested it is not an orthologue of the *Knotted-like* genes found in *A. thaliana*, which show a similar floral phenotype to that seen in *Oakleaf*. (Lincoln *et al.*, 1994; Serikawa *et al.*, 1996; Belles-Boix *et al.*, 2006; Li *et al.*, 2011; Cocker *et al.* 2015).

While these mapping data were able to provide some evidence to the position of the *S* locus, their applications were limited and their use alone would not lead to the discovery of the genes at the *S* locus. As a result, a search for molecular markers linked to the *S* locus was chosen as the best course of action in terms of physically locating the *S* locus.

1.6 Molecular markers for the *P. vulgaris* *S* locus

The three floral homeotic mutants discussed in section 1.5 provided a number of mapping distances for the *S* locus, however, further markers were needed to refine the position of the *S* locus, and these were obtained via molecular techniques.

1.6.1 *SLP1*

SLP1 was identified via Random Amplification of Polymorphic DNA (RAPD) analysis, in which polymerase chain reaction (PCR) experiments conducted on both pin and thrum genomic DNA resulted in a product that was only amplified when thrum DNA was used as the template (Manfield *et al.* 2005). The sequence of this product was used to produce a Sequence-Characterised Amplified Region (SCAR) marker that was able to determine whether DNA had been obtained from pin or thrum *P. vulgaris* plants. This marker was found to identify a Restriction Fragment Length Polymorphism (RFLP) using genomic DNA from the ‘Blue Jeans’ horticultural variety of *P. vulgaris* which had

identified a thrum-specific marker (Manfield *et al.* 2005). A *P. vulgaris* genomic DNA lambda phage library was screened with the RFLP marker *SLP*, and from this an 8.8 Kb contiguous sequence was constructed. It was, however, discovered that *SLP* was not present in the coding region of a gene, and markers that were actually expressed were sought (Li *et al.* 2007).

1.6.2 *SLL1* and *SLL2*

In the search for markers that were expressed in *P. vulgaris* tissues, two genes were identified: *SLL1* and *SLL2*. Fluorescent Differential Display (FDD) analysis was conducted on floral RNA from both pins and thrums. From this, 19 genes were isolated as showing differential expression between pin and thrum, however RT-PCR and Northern blotting resulted in only one of these genes showing differential expression. When tested via Southern analysis, two genes were shown to be linked to the *S* locus via RFLP analysis, and these were named *SLL1* and *SLL2* (Li *et al.* 2007). The *A. thaliana* gene similarity comparisons have suggested that *SLL1* is a small trans-membrane protein, whereas *SLL2* is suggested to be a *CONSTANS-LIKE* gene (Li *et al.* 2007).

Multiple alleles were found for both of the genes; two pin, one thrum. To obtain mapping distances from these genes to the *S* locus crosses were performed involving *P. vulgaris* individuals with known *SLL1* and *SLL2* alleles, and also including long and short homostyle plants, in which recombination events were proposed to have taken place within the *S* locus. The resultant data suggested that no recombination events occurred between *SLL1* and the *S* locus, its mapping distance was < 0.57 cM, and crosses using the long and short homostyles suggested that it was on the A side of the *S* locus. *SLL2* exhibited recombination events between it and the *S* locus, though rarely. It was shown to have a mapping distance of 1.57 cM from the *S* locus, and crosses using long and short homostyles suggested it was on the G side of the *S* locus (Li *et al.* 2007).

1.6.3 *GLO*

As discussed in section 1.5.1, the linkage of *GLO* to the *S* locus was utilised before the gene itself was discovered, by way of the overexpression mutant *Hose-in-Hose*. Three-point crosses involving this visible phenotype helped to shape the map that has been built of the area surrounding the *S* locus, however, the discovery of the gene responsible has further advanced the study of the *P. vulgaris* *S* locus.

In the search for the gene responsible for *sepaloid*, which presents as a loss-of-function B function mutant, the *P. vulgaris* orthologues of the *A. majus* B function genes *GLO* and *DEF* were discovered (Li *et al.* 2008). Expression of both genes was affected in the *sepaloid* mutant of *P. vulgaris*; however only *GLO* displayed linkage to the *S* locus. *GLO* was found to have three alleles, two pin and one thrum. It was suggested that *GLO* could not be the gene responsible for the *sepaloid* mutation as a plant showing the *sepaloid* mutation was found to have the thrum allele of *GLO*, and *sepaloid* had previously been linked to the pin floral form. Analysis of controlled crosses involving *P. vulgaris* pin and thrum individuals suggested that *sepaloid* was situated between *GLO* and the *S* locus (Li *et al.* 2008). It has since been found that the thrum allele of *GLO*, *GLO^T*, is spatially separated from *GLO* and unrelated to the *sepaloid* mutant, and is actually the result of a gene duplication event; the significance of this is discussed in section 1.9.1.

1.7 Genetic map of the *S* locus

The relationship of the *P. vulgaris* *S* locus to the genetic markers that had thus far been identified is shown in figure 1.6. The region in which the *S* locus is present had been identified, however this work has been entirely based on recombination, and as a result has not given insight into the structure of the *S* locus itself. For the mechanisms behind floral heteromorphy to be explained, the genes responsible must be identified and characterised.

Informed by the information obtained from recombinational analyses, two Bacterial Artificial Chromosome (BAC) libraries were created; a *Hind*II digestion, and a random shear library, both using DNA from the Blue Jeans horticultural cultivar of *P. vulgaris* (Li *et al.* 2011).

The four genetic markers identified above, the non-coding SLP1 and the *S* locus-linked genes *SLL1*, *SLL2* and *GLO* were used to identify BAC clones in which the regions flanking and containing the *S* locus would be included. This resulted in the discovery of 14 new *S* locus-linked genes, and over 2.2 Mb of genomic sequence surrounding the *S* locus, however the sequence across the *S* locus was not contiguous (Li *et al.* 2011). It was not possible to place one BAC contig within the map of the regions flanking the *S* locus, as it appeared to show no overlapping regions with any of the other contigs; it was possible that this contig was within the *S* locus itself, as contained within it was a putative gene, named *GLO^T*, which was present only on thrums. This gene was spatially separated from *GLO*, and was not merely a thrum-specific allele of *GLO*, as had been previously suggested (Li *et al.* 2007); this similarity will be discussed in detail in section 1.9.1. Despite this discovery, the BAC-walking method was unable to provide a contiguous coverage of the *S* locus, and as a result further methods had to be employed.

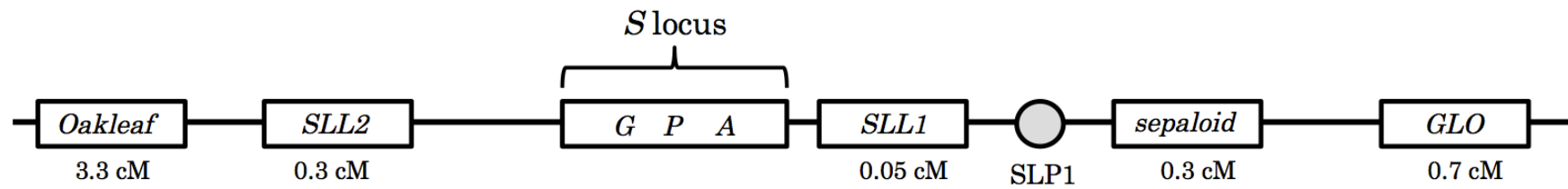


Figure 1.6: The *P. vulgaris* *S* locus in relation to known markers. Gene markers are indicated by rectangles; the non-coding marker *SLP1* is represented by a circle. Known linkage distances from the *S* locus are provided in centimorgans (cM); map is not shown to scale. Image modified from Smith 2014.

1.8 Genomic sequencing of wild type *P. vulgaris*

With the advent of Next Generation Sequencing (NGS) technologies came the ability to sequence entire genomes accurately and to an acceptable coverage depth. The vast majority of previous experiments to this date had been based on the Blue Jeans horticultural cultivar of *P. vulgaris*; which, as commercial cultivar, had been crossed with other *Primula* species in its creation.

The genome coverage levels achievable by the Illumina NGS technology are at their optimum when an inbred homozygous line is used as the starting material, and as a result, *de novo* genome sequencing of *P. vulgaris* was initiated by the Gilmartin group (Li *et al.* 2016), using a homozygous self-fertile long homostyle displaying the genotype *gPA/gPA*, plants of which originated from the long homostyle population previously found to exist in Somerset (Crosby 1940). Due to the assembly method of Illumina genome sequencing, the different alleles of the *S* locus could lead to the genome misassembling; the use of a highly homozygous inbred line was able to provide a scaffold to which the pin, thrum and short homostyle genomes could be aligned and compared, as the *de novo* sequence of pin and thrum genomes contain a much larger range of genetic diversity due to their obligate outcrossing nature (Li *et al.* 2015).

On the creation of this *de novo* genome assembly, the thrum-specific *GLO^T* sequence that had been identified as part of the BAC walk was searched against the long homostyle genome assembly, and this allowed two genome sequence contigs to be linked. This step facilitated the assembly of a 456 kb contiguous sequence which contained the *S* locus.

This region was compared to the genome sequence of the pin, thrum and short homostyle *P. vulgaris*, and in doing so it became apparent that a 278 kb region is absent from the pin genome (Li *et al.* 2016). While the sequence was found to be present in the genomes of the short homostyle and thrum plants, there appeared to only be one copy. This information was ascertained by assessing the average read-depth coverage of the genome sequence in the flanking regions of the *S* locus of the long homostyle, short homostyle, pin and thrum genomes. In all cases, it was found to be approximately 60X. Across the identified 278 kb region the coverage of the long homostyle genome remains at 60X, whereas that of the short homostyle and thrum genomes drops to approximately 30X (Li *et al.* 2016). This would suggest that long homostyle plants are homozygous for this region (*S/S*), short homostyle and thrum are hemizygous (*S/s*) and pins lack it entirely (*s/s*).

In the long homostyle, short homostyle and thrum genomes this insert is flanked on either side by two near-identical regions of approximately 3 kb, each of which contains *CYCLIN-LIKE F-BOX (CFB)* gene; in the pin genome, only one copy of the repeat is present. To both the left and right of the *CFB* genes that flank the *S* locus, a number of further flanking genes have been identified; these have shown to be present in both pin and thrum, however on inspection these regions have become homogenised by recombination, and are not part of the *S* locus themselves (Li *et al.* 2016). In figure 1.7 the *S* haplotype, in which the 278 kb region is present, is shown in comparison to the *s* haplotype, in which it is absent.

No further hemizygous regions have been found in the *P. vulgaris* genome (Li *et al.* 2016), and this 278 kb region and flanking sequences have also been shown to be present in the genomes of other *Primula* species, including *P. veris* and *P. elatior*, the two most closely related *Primula* species to *P. vulgaris* (Li *et al.* 2016). A hemizygous region behaves, genetically, as a homozygous one, however the implications of this are that the long and short homostyles did not arise by recombination, as previously proposed (Dowrick 1956). The genes found within this hemizygous region, and what is known about them, are both discussed in section 1.9.

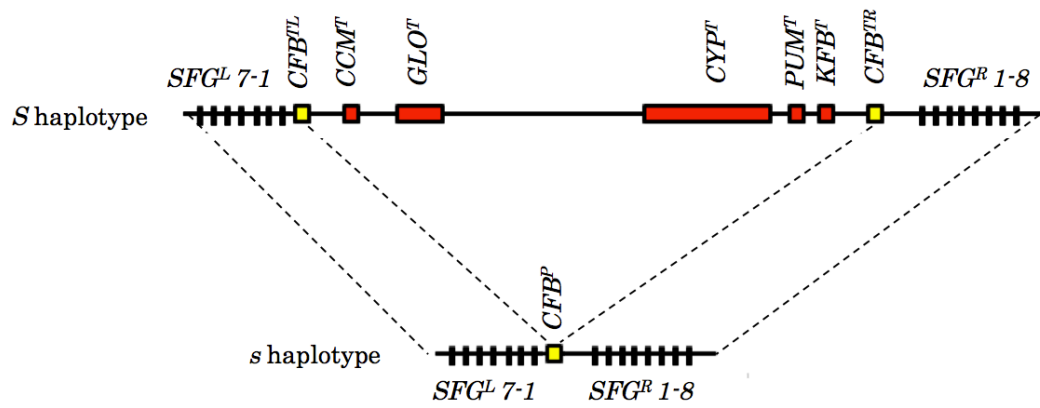


Figure 1.7: Comparison of the *S* and *s* haplotypes of the *S* locus. The *S* haplotype is specific to thrums. The *S* locus flanking genes to the left (*SFG^L*) and right (*SFG^R*) of the *S* locus are common to both haplotypes. The genes present within the thrum-specific *S* locus region are marked. Image modified from Li *et al.* 2016.

1.9 The genes within the *P. vulgaris* *S* locus

Within this region absent from the pin genome but present in the thrum, long homostyle and short homostyle there were found to be five genes; *CONSERVED CYSTEINE MOTIF* (*CCM^T*), *GLOBOSA* (*GLO^T*), *CYTOCHROME P450* (*CYP^T*), *PUMILIO-LIKE* (*PUM^T*), and *KISS-ME-DEADLY KELCH REPEAT F-BOX* (*KFB^T*). To date, little is known of the function of these genes, however, the identification of this region, and these genes, has led to the revelation that the recombination events that were proposed by Dowrick (1956) to have taken place to bring about the short and long homostyles did not occur, and Ernst's (1936b) original supposition that long and short homostyles had resulted from mutations was correct.

Work by the Gilmartin group has identified that the two long homostyle populations identified by Crosby (1940, 1949) both show a mutation in *CYP^T*, resulting in the elongation of the style. A recent publication (Huu *et al.* 2016) has identified the *P. veris* orthologue of *CYP^T*, named *CYP734A50*, which encodes a brassinosteroid-degrading enzyme. Quantitative real time PCR experiments have shown expression of this gene to be present predominantly in the style tissue, with minimal expression in the petal. In this study it has been suggested that a mutation in this gene is responsible for the elongated style seen in long homostyle plants (Huu *et al.* 2016), mirroring what has been found in *P. vulgaris*.

Genome sequencing conducted by the Gilmartin group on short homostyle plant DNA has shown that a transposon insertion into the coding region of *GLO^T* results in the gene not being expressed (Li *et al.* 2016). *GLO^T* was the first of the genes present in the thrum-specific region to be identified as being *S* locus linked, and resulted in the identification of the *S* locus (Li *et al.* 2016). As shown by the short homostyle phenotype, its role in the normal functioning of the *S* locus is key, as when knocked out the position of the anthers is altered, and this marked it as a candidate for performing one of the key functions required for heterostyly.

1.9.1 The duplication of *P. vulgaris* *GLO*

As mentioned in section 1.5, it was only recognised that *GLO^T* was a unique gene, and not a thrum-specific allele of the *P. vulgaris* floral development gene *GLO* when the two genes were shown to be spatially separated (Li *et al.* 2007). The two genes display a high level of sequence similarity, showing 82% identity at the protein level, and the gene duplication event is proposed to have occurred 52.7 million years ago

(Li *et al.* 2016).

Despite the high level of sequence identity between the two MADS box proteins, it has been shown that in thrum plants carrying the *Hose-in-Hose* mutation the expression of *GLO^T* is unaffected (Li *et al.* 2016). This would suggest that the duplicated gene could be performing a different function to that of *GLO*, and when combined with the phenotype shown when the expression of this gene is knocked-out, warrants further investigation.

GLO^T has not only been found to be present in the thrum, long homostyle and short homostyle genomes, but semi-quantitative PCR experiments conducted by Dr. Jinhong Li have shown that *GLO^T* is expressed in thrum and long homostyle flower bud tissue. That *GLO^T* is expressed in the flower bud tissue of these two genotypes, the floral morphologies of which display the anthers at the corolla mouth, whereas the short homostyle, in which it is not expressed shows low anthers, marks it as a likely candidate for the A function of heterostyly, raising the anthers in the thrum flower form.

1.10 Experimental aims

The presence of *GLO^T* in the thrum-specific hemizygous region of the *P. vulgaris* genome marks it as an interesting candidate for study, and the preliminary data into the expression of this gene suggests it is confined to the second and third whorl of the floral tissues. The absence of *GLO^T* expression from the short homostyle indicates it is playing a vital role in heterostyly. Little is yet known about *GLO^T*, and for its function to be elucidated more must be discovered about its localisation and expression. All of the information thus far discussed regarding *GLO^T* came about during the course of this investigation; at the start of this project, nothing was known other than its expression in the second and third whorl.

The *P. vulgaris* gene *GLO*, from which *GLO^T* is duplicated, has a defined function, specifying the organ identity of the second and third whorls, and its orthologues, *GLO* in *A. majus* and *PI* in *A. thaliana* have received extensive study. The extent to which *GLO^T* has diverged from its paralogue, in terms of expression patterns, localization, promoter sequence, and function will be analysed in this thesis.

In this investigation the temporal expression of the *P. vulgaris* genes *GLO* and *GLO^T* will be analysed across the floral developmental stages by way of qPCR, and the differences between the two paralogues will be compared. As no standard protocol or

validated reference genes are available for *P. vulgaris*, reference gene candidates will be selected and rigorous testing will be performed, to ensure accurate results.

The localisation of *P. vulgaris GLO* expression has previously been performed, however at the time the existence of *GLO^T* as a separate gene was not known. It is possible that cross-hybridisation was taking place between these two genes. The localisation patterns of *GLO* will be confirmed, and that of *GLO^T* will be ascertained. Localisation experiments for the *GLO* and *GLO^T* proteins will be planned, and peptide antibodies will be designed that will specifically recognize the protein to which they are designed.

The promoter sequences of *P. vulgaris GLO* and *GLO^T* will be compared, and the different MADS box transcription factor binding sites that are found within them will be analysed, as this will provide information as to what is controlling the expression of these genes. Promoter-reporter constructs will be designed and generated for use in *A. thaliana* with the aim of assessing the localisation patterns of *GLO* and *GLO^T*.

Yeast 2-Hybrid (Y2H) experiments will be conducted to ascertain whether protein-protein interactions are present between the *A. thaliana* B function proteins *PI* and *AP3* and the *P. vulgaris* orthologues, *GLO* and *DEF*, and *GLO^T*, and information from these experiments will be used to inform *A. thaliana* complementation experiments. The *pi-1* mutant in *A. thaliana* can be complemented when the native *PI* gene has been reintroduced, and the introduction of a *PI* orthologue from another species can also result in complementation. As such, constructs will be designed and generated in order to determine whether *GLO* and *GLO^T* are able to complement the *pi-1* mutant.

Finally, a *GLO^T* overexpression construct has been generated and introduced into *A. thaliana* by Dr. Sadiye Hayta. The production of the equivalent construct for *GLO* will be performed, with the intention of comparing the overexpression phenotypes of *GLO* and *GLO^T*.

CHAPTER 2

Materials and methods

2.1 List of suppliers

Bio-Rad Laboratories Ltd.

(www.bio-rad.com)

Bio-Rad House, Maxted Road, Hemel Hempstead, HP2 7DX, United Kingdom

Biotechnology Resources for Arable Crop Transformation (BRACt)

(<http://www.bract.org/>)

John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

Eurofins MWG Operon

(www.mwg-biotech.com)

Anzinger Straße 7a 85560, Ebersberg, Germany

GE Healthcare Life Sciences

(<http://www.gelifesciences.com/>)

Amersham Place, Little Chalfont, Buckinghamshire, HP7 9NA, United Kingdom

Leica Microsystems (UK) Ltd

(<http://www.leicabiosystems.com/>)

Larch House, Woodlands Business Park, Breckland, Linford Wood, Milton Keynes,
MK14 6FG, United Kingdom

Nottingham Arabidopsis Stock Centre

(<http://arabidopsis.info>)

School of Biosciences, University of Nottingham, Sutton Bonington Campus,
Loughborough, LE12 5RD, United Kingdom

New England Biolabs Ltd.

(www.neb.uk.com)

5-77 Knowl Piece, Wilbury Way, Hitchin, SG4 0TY, United Kingdom

Promega UK Ltd.

(www.promega.com)

Enterprise Rd, Southampton, SO16 7NS, United Kingdom

QIAGEN Ltd.

(www.qiagen.com/)

Skelton House, Lloyd Street North, Manchester, M15 6SH, United Kingdom

Roche Diagnostics Ltd.

(www.roche.co.uk/portal/uk/researchers_)

Charles Avenue, Burgess Hill, RH15 9RY, United Kingdom

Sakura Finetek UK Ltd.

(<http://www.sakura.eu/home>)

1 Thatcham Business Village, Colthrop Way, Thatcham, RG19 4LW, United
Kingdom

Sigma-Aldrich Company Ltd.

(www.sigmaaldrich.com/united-kingdom.html)

The Old Brickyard, New Road, Gillingham, SP8 4XT, United Kingdom

Takara Bio Europe

(<http://www.clontech.com>)

2 Avenue du Président Kennedy, 78100 Saint-Germain-en-Laye, France

Thermo Fisher Scientific Ltd. (UK)

(<http://www.thermoscientific.com>)

Stafford House, Boundary Way, Hemel Hempstead, HP2 7GE, United Kingdom

VWR International Ltd.

(<https://uk.vwr.com/store>)

Hunter Boulevard, Magna Park, Lutterworth, Leicestershire, LE17 4XN, United Kingdom

2.2 Stock solutions and media

2.2.1 Chemicals

All chemicals were supplied by Sigma-Aldrich Company Ltd. unless otherwise specified.

2.2.2 Stock solutions

0.5 M EDTA:	pH was adjusted with NaOH to 8.0. Autoclaved.
1 M MgCl₂:	Autoclaved.
5 M NaCl:	Autoclaved.
1 M Tris-HCl:	pH was adjusted with HCl before autoclaving. pH 7.0, 7.5, 8.0 and 9.5 were all used.
4M LiCl:	Autoclaved.
100 mM Triethanolamine:	Added to ddH ₂ O, adjusted to pH 8.0, made immediately before use.
10% Glycine (w/v):	Dissolved in water, autoclaved.

2.2.3 Stock buffers

Edwards Buffer:	250 mM NaCl, 200 mM Tris-HCl (pH 7.5), 25 mM EDTA (pH 8.0), 0.5% SDS (v/v)
2X CO₃:	80 mM NaHCO ₃ , 120 mM Na ₂ CO ₃ Autoclaved before use.
10X TBE:	1 M Tris, 1 M H ₃ BO ₃ , 0.02 M EDTA (pH 8.0)
20X SSPE:	3 M NaCl, 20 mM EDTA, 200 mM NaH ₂ PO ₄ Adjusted to pH 7.4, autoclaved before use.
10X PBS:	1.3 M NaCl, 70 mM Na ₂ HPO ₄ , 30 mM NaH ₂ PO ₄ , 2.7 mM KCl Adjusted to pH 7.4, autoclaved before use.
10X TE:	100 mM Tris-HCl (of desired pH), 10 mM EDTA, (pH 8.0) Autoclaved before use.
10X NTE:	5 M NaCl, 100 mM Tris-HCl (pH 7.5), 10 mM EDTA (pH 8.0) Tris-HCl and EDTA added from liquid stocks, autoclaved before use.
50X Denhardt's solution:	Ficoll 400 1% (w/v), Polyvinylpyrrolidone (PVP) 1% (w/v), Bovine serum albumin (Fraction V) 1% (w/v) dissolved in 100 ml deionised water. Filter sterilised before use.

2.2.4 Growth media

Luria-Bertani (LB) medium:	10 g NaCl, 10 g tryptone, 5 g yeast extract, in 1 L of deionised water. Adjusted to pH 7.0 with 1 M NaOH and autoclaved.
LB agar:	LB medium (as above) containing 1.5% (w/v) Lab M agar. Autoclaved.
Super Optimal broth with Catabolite repression (SOC):	20 g tryptone, 5 g yeast extract, 0.58 g NaCl, 0.19 g KCl, 2.03 g MgCl ₂ , 2.46 g MgSO ₄ •7H ₂ O, 3.6 g glucose in 1 L of deionised water. Autoclaved.
1X Murashige and Skoog (MS) + 0.8% agar:	4.41 g MS medium (including vitamins), 30 g sucrose in 1 L of deionised water. Adjusted to pH 5.8 with 1 M NaOH, added 0.8% (v/w) Formedium agar and autoclaved.

2.3 Plant material

Plants and seeds of *Primula vulgaris* and *Primula veris* were obtained from the population maintained by the Gilmartin lab at the University of East Anglia. Leaf and flower bud tissue of *Primula elatior* was obtained from adult plants found in Bull Wood (Bury St. Edmunds, UK) with permission given by the Suffolk Wildlife Trust. All mature plants were maintained in greenhouses at the University of East Anglia.

Arabidopsis thaliana seeds of the Landsberg *Erecta* and Columbia-0 ecotypes were obtained from the JIC seed store. *A. thaliana* seeds for use in the complementation of the *pistillata-1* mutant were obtained from the Nottingham Arabidopsis Stock Centre; the supplied seed were generated by the self-pollination of Ler/*pi-1* plants.

2.4 Seed sterilisation and germination

2.4.1 *Primula* species

Seeds were collected from all *Primula* species and dried thoroughly. Seeds were sterilised by immersion in 70% ethanol (v/v) for 2 minutes followed by a 15-minute incubation in 10% (v/v) sodium hypochlorite. Seeds were rinsed three times in distilled water and were incubated overnight in 400 ppm GA₃ to remove the stratification requirement. Seeds were germinated on 1X Murashige and Skoog (MS) + 0.8% agar. After germination, seeds were transferred to soil and were grown in the greenhouses at the University of East Anglia.

2.4.2 Standard *A. thaliana* seed sterilisation

Arabidopsis seeds were transferred to a 1.5 ml microcentrifuge tube and surface-sterilised; all steps were carried out in a sterile flow hood. 500 µl of 7.5% (v/v) sodium hypochlorite was added to each tube of seeds, tubes were inverted five times and all samples were incubated for no longer than 7 minutes. 500 µl of 0.1% (v/v) Triton-X 100 was added to each tube, samples were mixed, and the solution was pipetted off and discarded. 1 ml of 0.1% Triton-X 100 was added and the tubes were inverted. The liquid was pipetted off and 1 ml of autoclaved distilled water was added to each tube, tubes were inverted, and the water was removed. One ml of water was added to each tube and removed a further eight times. Seeds were

dispensed onto sterile 90 mm filter paper discs and were allowed to dry. When the filter paper discs were dry, the seeds were distributed onto agar plates containing 1X MS + 0.8% agar.

2.4.3 Sterilisation of transformed *A. thaliana* seeds

Seeds obtained following transformation via *Agrobacterium tumefaciens* were sterilised via an alternative method, as the process of dipping plants resulted in increased fungal contamination. To sterilise, seeds were dispensed into 1.5 ml microcentrifuge tubes, which were placed in a rack inside a sealable container in a fume hood. A flask was placed inside the sealable container, and to this 97 ml of 13% (v/v) sodium hypochlorite was dispensed. Three ml of 1 M HCl was added to the flask and the sealable container was closed immediately. Seeds were left sterilising overnight, and tubes were closed prior to removal from the fume hood. Seeds were then plated as described above.

2.4.4 *A. thaliana* growth conditions

In a laminar flow cabinet, sterilised seeds were sown onto 1X MS + 0.8% agar plates. These were stored at 4 °C for four days to allow them to stratify. After four days the agar plates were then transferred to constant light conditions at 25 °C. When seedlings were approximately 14 days old they were transferred to the John Innes Centre Arabidopsis mix (Levington F2 600 L peat, 100 L 4 mm grit, 196 g Exemptor [Chloronicotiny] Insecticide), and grown in the John Innes Centre greenhouses.

2.5 Nucleotide isolation

2.5.1 Genomic DNA extraction

Three methods of extraction were used to obtain genomic DNA from plant tissue throughout this investigation. The illustra Nucleon PhytoPure Extraction Kit (GE Healthcare Life Sciences) was used for plant samples above 1 g in weight. The standard protocol was followed. For smaller samples up to 0.1 g, the DNeasy Plant Mini Kit (QIAGEN Ltd.) was used, following the manufacturer's instructions. A modified version of the Edwards DNA Extraction method (Edwards *et al.* 1991) was also used on *A. thaliana* tissue for genotyping purposes only.

2.5.1.1 illustra Nucleon PhytoPure Extraction Kit

Pre-weighed plant tissue was transferred to a mortar and pestle pre-cooled with liquid nitrogen, and samples were ground until a free-flowing powder was produced. The powder was transferred to a centrifuge tube and 4.6 ml of Reagent 1 added, followed by RNase A to a final concentration of 20 µg/ml. Samples were incubated at 37 °C for 30 minutes. After the incubation step, 1.5 ml of Reagent 2 was added, and samples were inverted to obtain a homogenous mixture. Samples were heated to 65 °C for 10 minutes, with regular manual agitation throughout incubation. After heating samples were placed on ice for 20 minutes. ml of chloroform cooled to -20 °C was added to the sample to aid the removal of proteins and polysaccharides. 200 µl of Nucleon PhytoPure DNA Extraction Resin was added to the sample, which was manually agitated at regular intervals for 10 minutes, before centrifugation at 1,300 *g* for 10 minutes. After centrifugation, the upper aqueous phase was transferred to a 1.5 ml microcentrifuge tube, with care being taken to not disturb the resin layer. An equal volume of isopropanol was added to the sample, and the tube was gently inverted to precipitate the DNA. Samples were centrifuged at 4,000 *g* for 5 minutes to pellet the DNA, before the supernatant was discarded and the pellet briefly washed with 70% ethanol. The sample was once again centrifuged at 4,000 *g*, for 5 minutes to pellet the DNA. The supernatant was discarded and the pellet was air-dried for 10 minutes, before re-suspension in 1X TE buffer (pH 8.0) or water.

2.5.1.2 QIAGEN DNeasy Plant Mini Kit

A maximum of 100 mg of plant tissue was ground in a 1.5 ml microcentrifuge tube with a micropestle and 400 µl of Buffer AP1 and 4 µl of RNase A stock solution (100 mg/ml) added to the sample before vortexing to ensure distribution of the reagents. Samples were incubated at 65 °C for 10 minutes, with several inversions throughout the incubation step to lyse the cells and 130 µl of Buffer P3 added to the lysate to precipitate detergent, proteins and polysaccharides. The samples were then incubated on ice for 5 minutes, before centrifugation at 20,000 *g* for 5 minutes. The supernatant was transferred to a QIAshredder Mini spin column, and centrifuged for 2 minutes at 20,000 *g*. The flow-through was transferred to a clean 1.5 ml microcentrifuge tube, with care being taken not to disturb the cell-debris pellet. 1.5 sample volumes of Buffer AW1 were added to the lysate and the solutions were mixed by pipetting. 650 µl of the mixture was transferred to a DNeasy Mini spin column and was spun at 6,000 *g*; the process was repeated with any remaining

mixture, using the same spin column for the remainder of each sample. The spin column was transferred to a clean 2 ml collection tube, 500 µl Buffer AW2 was added, and the sample was spun for 1 minute at 6,000 *g*. The 500 µl Buffer AW2 wash was repeated, and was centrifuged for 2 minutes at 20,000 *g*. The spin column was then transferred to a clean 1.5 ml microcentrifuge tube, and 50 µl Buffer AE was pipetted directly onto the membrane of the spin column. The sample was incubated at room temperature for 5 minutes, before centrifugation at 20,000 *g* for 1 minute. A second elution of 50 µl was performed into a separate clean 1.5 ml microcentrifuge tube.

2.5.1.3 Edwards genomic DNA extraction

This is a modified version of the method detailed in Edwards *et al.* 1991. One rosette leaf from an adult *Arabidopsis* plant and was placed in a 1.5 ml microcentrifuge tube and immediately frozen in liquid nitrogen. The frozen tissue was ground using a micropestle, and 400 µl of Edwards Extraction Buffer was added to each sample, and the sample was vortexed briefly. Samples were collected at room temperature until all had been ground and buffer had been added. Samples were spun for 1 minute at 14,500 *g* in a benchtop microcentrifuge and 300 µl of the supernatant transferred to a fresh 1.5 ml microcentrifuge tube and 300 µl of isopropanol added to each sample. Samples were mixed by inversion and incubated at room temperature for 2 minutes. Samples were spun for 10 minutes at 14,500 *g* in a benchtop microcentrifuge and the supernatant was removed by pipetting. All samples were spun under vacuum for approximately 5 minutes in a vacuum concentrator to aid drying. 50 µl of 1X TE buffer (pH 8.0) was added to each sample, and samples were transferred to 4 °C and incubated overnight to allow the DNA to re-dissolve. Samples were transferred to -20 °C for long-term storage.

2.5.2 RNA isolation

RNA isolation was carried out using the RNeasy Total RNA Isolation Kit (Thermo Fisher Scientific Ltd.), following the manufacturer's instructions that specifically referred to plant RNA. To minimise tissue loss, tissue was ground within an RNase-free 1.5 ml microcentrifuge tube partially submerged in liquid nitrogen. A motorised micropestle was used to grind the tissue to a fine powder, on which the sample tube was removed from liquid nitrogen and transferred to ice. 10-12 volumes of Lysis/Binding Buffer and one volume of Plant RNA Isolation Aid were added to the disrupted tissue. Samples were vortexed thoroughly and collected on ice until all samples had been processed. Samples were centrifuged at 14,500 *g* to pellet cell debris. Supernatant was transferred to a fresh 1.5 ml microcentrifuge tube and an equal volume of 64% ethanol was added; samples were mixed by vortexing. The mixture was applied to a filter cartridge that had been placed in a collection tube, and samples were centrifuged at 20,000 *g* for 1 minute. The flow through was discarded and the previous spin repeated until all lysate had passed through the filter cartridge before 700 µl of Wash Solution 1 was applied directly to the membrane of the filter cartridge. Samples were spun at 20,000 *g* and the flow through was discarded. 500 µl of Wash Solution 2/3 was applied to the membrane. Samples were once again spun at 20,000 *g* and the flow through was discarded. A second 500 µl of Wash Solution 2/3 was applied to the membrane and a further 1-minute spin at 20,000 *g* was conducted. The filter cartridge and collection tube were centrifuged for 1 minute at 20,000 *g* to remove any residual buffer, and the filter cartridge was transferred to a clean 1.5 ml microcentrifuge tube. 40 µl of Elution Solution pre-heated to 80 °C was applied to the membrane. The sample was spun for 1 minute at 20,000 *g*, before a further 20 µl of Elution Solution was added and the spin repeated. RNA concentration was verified using a Qubit Fluorometer 2.0 (Thermo Fisher Scientific Ltd.), and 260/280 and 260/230 ratios were checked using Nanodrop 1000 to identify potential contaminants.

2.5.3 Precipitation of RNA

RNA was precipitated to remove contaminants and to increase the concentration of samples prior to cDNA synthesis; 3 M sodium acetate (Thermo Fisher Scientific Ltd.) was used as the precipitating salt. 10 µl of 3 M sodium acetate was added for every 100 µl of RNA sample. To this, 3 volumes of pre-cooled 100% ethanol were added. Samples were mixed well, and incubated at -80 °C for a minimum of 1 hour. Samples were spun at 20,000 *g* in a centrifuge capable of maintaining a temperature of 4 °C for 20 minutes. The supernatant was removed, and samples were washed with pre-chilled 70% ethanol. Samples were spun again at 20,000 *g* in a centrifuge capable of maintaining a temperature of 4 °C for 10 minutes. The 70% ethanol was removed, and samples were allowed to air dry on the benchtop. Samples were re-suspended in nuclease-free water of the desired volume.

2.5.4 cDNA synthesis

cDNA synthesis for conventional PCR was conducted using M-MLV and oligoDT. For quantitative real time PCR a different method was employed, using random primers. Random primers also give higher yields of cDNA, and as the expression of the genes at the *S* locus is known to be low, maximising the yield was important. One added advantage of random priming is that it generates the least bias in the resulting cDNA (Ginzinger 2002).

2.5.5 DNase treatment of RNA

All RNA was treated with RQ1 RNase-Free DNase (Promega UK Ltd.) before cDNA synthesis. Between 1 and 8 µl of RNA was transferred to a 0.2 ml tube for each sample, and 1 µl of RQ1 RNase-Free DNase 10X Reaction Buffer was added. For every 1 µg of RNA, 1 µl of RNase-Free DNase was added, and the reaction was made up to 10 µl with nuclease-free water. Samples were incubated at 37 °C for 30 minutes. One µl of RQ1 DNase Stop Solution was added to each sample, and samples were incubated at 65 °C for 10 minutes to inactivate the DNase.

2.5.6 cDNA synthesis for standard PCR

M-MLV Reverse Transcriptase (Thermo Fisher Scientific Ltd.) was used to transcribe the RNA into cDNA. 1 µl of 500 µg/ml oligoDT (Promega UK Ltd.) and 1 µl of 10 mM dNTPs (Thermo Fisher Scientific Ltd.) were added to 10 µl of DNase-treated RNA. Samples were incubated at 65 °C for 5 minutes and chilled on ice. 4 µl of 5X First Strand Buffer, 2 µl of 0.1 M DDT and 1 µl of RNase OUT (all supplied by Thermo Fisher Scientific Ltd.) were added to each reaction and samples were incubated at 37 °C for 2 minutes. One µl of M-MLV Reverse Transcriptase was added to each sample, and all samples were incubated at 37 °C for 50 minutes. The reaction was inactivated by incubation at 70 °C for 15 minutes. Complementary RNA was removed via digestion with RNase H (Promega UK Ltd.); 1 µl was added to each sample and all samples were incubated at 37 °C for 15 minutes. The reaction was terminated by a 15-minute 65 °C incubation.

2.5.7 cDNA synthesis for quantitative real-time PCR

The High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific Ltd.) was used to synthesise cDNA for qPCR. The amount of RNA in each reaction was standardised; DNase-treated RNA was quantified using the Qubit Fluorometer 2.0.

All samples were diluted to 2 µg in 10 µl. To each sample 2 µl of 10X RT Buffer, 0.8 µl of 25X dNTP Mix (100 mM), 2 µl of 10X RT Random Primers, 1 µl of Multiscribe Reverse Transcriptase, and 4 µl of nuclease-free water were added. Each set of samples included a 'No Template Control' in which water was substituted for RNA and a 'No Reverse Transcriptase' in which water was substituted for Multiscribe Reverse Transcriptase to control for genomic DNA contamination. Samples were vortexed and centrifuged briefly before transferral to a T100 Thermal Cycler (Bio-Rad Laboratories Ltd.). The following cycling program was used: 25 °C for 10 minutes, 37 °C for 120 minutes and 85 °C for 5 seconds. Samples were stored at -20 °C until use.

2.6 Polymerase Chain Reaction

2.6.1 Oligonucleotide design and synthesis

Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>) was used to design oligonucleotides primers for all PCR reactions conducted as part of this investigation. All oligonucleotides were ordered from Eurofins MWG Operon.

2.6.2 Standard PCR

GoTaq Flexi polymerase (Promega UK Ltd.) is a Taq polymerase lacking 3' to 5' exonuclease proofreading capability, and was used to amplify cDNA and genomic DNA fragments between 100 bp and 1.5 kb in size. The supplied kit contains all reagents other than dNTPs, which were obtained from Thermo Fisher Scientific Ltd.

Each 50 µl reaction contained: 1X Green GoTaq Flexi Buffer, 2 mM MgCl₂, 1 mM of each primer, 0.2 mM of each dNTP, 1.25 units of TAQ polymerase, and between 1-5 µl of DNA; DNA volume was dependant on template concentration, with 100 ng of gDNA considered the optimum for PCR. A standard GoTaq PCR consisted of a 2-minute initial denaturation step at 95 °C, followed by between 20-35 cycles of denaturation at 95 °C for 30 seconds, annealing at between 50 and 65 °C for 30 seconds and an extension step at 72 °C for 1 minute. A final extension step of 5 minutes concluded the reaction, and samples were stored at 4 °C until use. All reactions were carried out using a T100 Thermal Cycler (Bio-Rad Laboratories Ltd.).

2.6.3 Long and proofreading PCR

When high-fidelity PCR reactions were required or when a fragment was larger than 1.5 kb in size Phusion High-Fidelity DNA Polymerase (New England Biolabs Ltd.) was used, as it possesses the ability to conduct 3' to 5' exonuclease proofreading. The supplied kit contains all reagents other than dNTPs, which were obtained from Thermo Fisher Scientific Ltd.

Each 50 µl reaction contained, 0.5 µM of each primer, 10 µl of 5X HF Phusion buffer, 200 mM dNTPs, 0.01 units of Phusion, and between 1-5 µl of DNA, with an optimum of 100 ng. The typical cycling conditions consisted of a 30 second initial denaturation step at 98 °C, followed by between 25-35 cycles of denaturation at 98 °C for ten seconds, annealing at between 55 and 65 °C for 30 seconds and extension at 72 °C

for 30 seconds, with a final extension step at 72 °C for 10 minutes. Samples were stored at 4 °C until use, or at -20 °C for long-term storage. All reactions were carried out using a T100 Thermal Cycler (Bio-Rad Laboratories Ltd.)

2.6.4 Gel Electrophoresis

Nucleic acids were quantified by agarose gel electrophoresis on 0.8-2% (w/v) agarose gels; the percentage chosen was dependent upon the size of the fragment being analysed. Agarose was melted in 0.5X TBE, and was allowed to cool to approximately 50 °C before the addition of ethidium bromide at a final concentration of 0.2 µl/ml. The warm gel was poured into a cast and allowed to set, before submersion in 0.5X TBE in a gel electrophoresis tank. For samples that did not have loading dye included in the PCR buffer, DNA loading dye (New England Biolabs Ltd.) was added (2.5% Ficoll®-400, 11 mM EDTA, 3.3 mM Tris-HCl, 0.017% SDS, 0.015% bromophenol blue). Alongside PCR samples, a DNA molecular weight marker was run; 100 bp or 1 kb ladder (New England Biolabs Ltd.) was chosen depending upon the DNA fragment being analysed. Gels were run at 60 V for 40-60 minutes, and were visualised using a UV transilluminator.

2.6.6 Isolation of DNA from PCR products

Fragments were excised from the agarose gel under UV light, using a clean scalpel and transferred to a microcentrifuge tube, which was then weighed and the weight of the sample recorded. If a PCR reaction resulted in multiple bands being present on the gel, or a restriction digest had been performed and only one fragment was required, the QIAquick Gel Extraction Kit (QIAGEN Ltd.) was used. If a PCR reaction resulted in a single band when run on a gel, the QIAquick PCR Purification Kit (QIAGEN Ltd.) was used to extract the remainder of the PCR product; this method did not require the sample to be run on an agarose gel, hence increasing the yield of the product.

2.6.7 QIAquick Gel Extraction Kit

Three volumes of the gel solubilisation buffer Buffer QG were added per volume of gel, with 1 mg of DNA-containing gel corresponding to 1 µl of buffer. The purpose of Buffer QG, which contains guanidine thiosulphate, is to remove elements in the PCR reaction other than the single or double stranded DNA, such as the polymerase. Gel samples were incubated in a heat block at 50 °C for 10 minutes. One gel volume of isopropanol was added to the sample to increase the yield of DNA, and each sample was transferred to a QIAquick spin column and collection tube, and spun at 14,500 *g* in a benchtop centrifuge for 1 minute. A second wash of Buffer QG was conducted, and 500 µl of QG was added to the column, and the spin was repeated to remove all traces of agarose before 750 µl of Buffer PE was added to wash the sample and prevent dissociation of DNA from the column membrane. This was allowed to stand for between 2 and 5 minutes to ensure DNA was retained before centrifuging for 1 minute. The flow-through was then discarded and the sample re-spun to remove any remaining ethanol. DNA was eluted into 30 µl of Buffer EB (10 mM Tris-HCl [pH 8.0]) to maximise DNA concentration, and up to 4 minutes of standing time was allowed to increase DNA binding. The sample was then spun for 1 minute and eluted DNA was checked using a Qubit Fluorometer 2.0.

2.6.8 QIAquick PCR Purification Kit

The PCR purification kit follows the same principles as the Gel extraction kit, however five volumes of Buffer PB are used in place of three volumes of QG. This buffer contains the guanidine hydrochloride to allow DNA to bind to the membrane of the spin column. After this substitution, all other steps occur as described above.

2.6.9 DNA sequencing

All single read sequencing of PCR products and plasmids was carried out by Eurofins MWG Operon.

2.6.10 A-tailing of PCR products

DNA polymerases, such as GoTaq Polymerase, that lack the 3' – 5' proofreading capability of higher-fidelity enzymes add an adenine onto each end of the PCR product. This allows them to be combined with TA cloning systems with no need for digestion of PCR product or vector prior to ligation. However, these enzymes have a relatively high error rate (one error per 1000 bp) and as a result are not suitable for occasions when sequence accuracy is crucial. Proofreading enzymes, such as Phusion, have a much lower error rate, but do not add this extra adenine and as a result cannot be ligated directly into TA cloning vectors.

To ensure PCR products produced by proofreading enzymes are compatible with TA cloning vectors this extra adenine must be added onto the end of the PCR product. PCR products were purified using the QIAquick PCR Purification Kit, and 4 µl of this cleaned PCR product was combined with 2 µl of 5X GoTaq Reaction Buffer, 2 µl of 1 mM dATP (0.2 mM final concentration), 1 µl GoTaq Flexi DNA Polymerase, and 0.6 µl 25 mM MgCl₂ (1.5 mM final concentration); samples were then made up to 10 µl with nuclease-free water. Samples were incubated at 70 °C for 30 minutes to facilitate the addition of this extra adenine. 2 µl of this was then used in a pGEM-T Easy ligation reaction (Maciver 2010).

2.7 Quantification of transcript by quantitative real-time PCR (qPCR)

In conventional PCR the amount of product produced is measured at the end-point of the reaction when it has reached saturation, and as a result no conclusions can be drawn from comparing one sample to the next. In qPCR, the amount of product is measured at the end of every PCR cycle, allowing samples to be compared to one another. When the fluorescent SYBR Green I dye is used, it only emits light when bound to double-stranded DNA. As the amount of PCR product increases, the strength of the fluorescent signal emitted by the SYBR also increases.

2.7.1 The standard qPCR reaction

All 96-well qPCR reactions were conducted using a CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories Ltd.), all 384-well qPCR reactions were conducted using a LightCycler® 480 System (Roche Diagnostics Ltd.).

Reactions were set up to a final volume of 15 µl using SYBR Green JumpStart Taq ReadyMix kit (Sigma-Aldrich Company Ltd.). Before use, all cDNA samples were diluted to 1:40, quantified using a Qubit Fluorometer 2.0, and standardised to the weakest sample. This was done to counteract for any variations in the efficiency of the cDNA synthesis reaction. The reactions contained 7.5 µl of SYBR, 0.9 µl of each primer (final concentration 400 nM) and 0.7 µl of water with and 5 µl of the diluted cDNA sample.

The same cycling program was used for all samples: 94 °C for 2 minutes. 94 °C for 15 seconds, 60 °C for 1 minute, repeated 40 times. All reactions were run with a melt curve post-cycling to check the specificity of the targets. This began with a 10 second 95 °C step, and was followed by a melt curve beginning at 65 °C and ending at 95 °C, increasing by an increment of 0.5 °C. Each step of the melt curve was held for 5 seconds. Data were analysed via the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

2.7.2 Testing of all primers for qPCR

A standard GoTaq Flexi polymerase PCR was conducted for all qPCR primers; all samples were run on an agarose gel to confirm that only one PCR product was produced. Primer combinations that did produce only one band were then subjected to further amplification efficiency checks.

All cDNA samples that each primer combination was to be tested on were pooled and diluted to 1:10, 1:20, 1:40, 1:80 and 1:160 of the original concentration. A standard qPCR reaction was then run for each of the primer pairs, testing each of the dilutions. The resultant C_q values were plotted against the dilution factor of the cDNA. Primers that are amplifying product efficiently will display an increase of one C_q unit when the cDNA concentration has halved.

2.7.3 Selection and testing of genes for normalisation of qPCR results

Genes used for normalisation require detailed testing, as a gene that does not normalise correctly can influence the results seen. A number of reference genes were tested; the process by which this occurred is described in detail in Chapter 3. A standard GoTaq PCR was conducted for every primer pair, and for primer pairs that produced one band; a cDNA dilution efficiency was conducted to check for accuracy, as was a melt peak test, both as described above.

2.8 Molecular cloning

2.8.1 Ligation reaction for pGEM-T Easy constructs

The ligation of fragments into pGEM-T Easy (Promega UK Ltd.) was always conducted at 4 °C overnight; this method yields the maximum number of recombinant molecules. The standard reaction consisted of 5 µl of 2X Rapid Ligation Buffer, 1 µl of the pGEM-T Easy Vector (50 ng), 1 µl T4 DNA Ligase and variable amounts of PCR product and water. The required amount of PCR product was determined by the following equation:

$$\frac{ng\ of\ vector\ \times\ kb\ size\ of\ insert}{kb\ size\ of\ vector} \times insert:vector\ molar\ ratio = ng\ of\ insert$$

All pGEM-T Easy constructs were propagated in DH5α competent cells prepared in the Gilmartin lab; the method is described in the subsequent section.

2.8.2 Preparation of DH5α competent cells

Competent cells were grown from a stock of sub-cloning efficiency DH5α competent cells (Thermo Fisher Scientific Ltd.). A clean pipette tip was touched lightly to the inside of the tube of cells obtained from the supplier. This tip was subsequently dragged lightly across the surface of an LB plate containing 100 mg ampicillin. The plate was incubated in a 37 °C oven overnight. Five ml of LB medium was inoculated with a single colony that had grown on the plate. This liquid culture was grown overnight in a shaking incubator (220 rpm) set to 37 °C. 1 ml of this overnight culture was used to inoculate 100 ml of LB culture medium. This was incubated at 37 °C at 220 rpm until the optical density (OD₆₀₀) of the culture measured between 0.3 and 0.6. At this point the culture was transferred to ice before being spun at 3,000 g for 5 minutes in a refrigerated centrifuge at 4 °C. Following centrifugation the supernatant was poured off and the pellet was re-suspended in 25 ml of ice cold 0.1 M MgCl₂. The culture was incubated on ice for 10 minutes and was then spun at 3,000 g at 4 °C. The pellet was then re-suspended in 25 ml of ice cold 0.1 M CaCl₂. The culture was incubated on ice for 10 minutes and spun at 3,000 g for 5 minutes at 4 °C. A 1.7 ml aliquot of 0.1 M CaCl₂ and 0.3 ml of glycerol were mixed, and 1.25

ml of this was used to re-suspend the pellet. These were then aliquoted, frozen immediately in liquid nitrogen, and stored at -80 °C until use.

2.8.3 Transformation of DH5α competent cells with pGEM-T Easy constructs

A 50 µl aliquot of DH5α competent cells prepared as described in 2.3.2 was thawed on ice, and 1 µl of ligation mixture described in 2.3.1 was added. The cells were then incubated on ice for 30 minutes. A 30-second heat shock was carried out at 42 °C before cells were returned to ice for a further 2 minutes. 700 µl of LB medium was added, and cells were allowed to grow for one hour in a 37 °C incubator, with shaking at 225 rpm. Cells were subsequently spun in a benchtop microcentrifuge at 3,000 *g* for 6 minutes. Excess media was removed and cells were re-suspended in 50 µl of fresh media. Cells were plated out onto LB agar containing 50 µg/ml X-Gal and 100 µg/ml ampicillin, under a Bunsen flame for sterility. Plates were transferred to a 37 °C incubator for overnight growth.

2.8.4 Ligation reaction for Gateway constructs

pCR8 (Thermo Fisher Scientific Ltd.) was used as the entry vector for all Gateway cloning applications. PCR products were extracted and A-tailed as described previously before ligation into pCR8.

The ligation reaction consisted of 4 µl of A-tailed PCR, 1 µl Salt Solution, and 1 µl pCR8. The reaction was gently mixed and incubated at room temperature for 30 minutes, before transformation of One Shot TOP10 competent cells (Thermo Fisher Scientific Ltd.), as described below.

2.8.5 Transformation of TOP10 competent cells with Gateway constructs

Two µl of the pCR8 ligation mix was added to 50 µl of One Shot TOP10 competent cells, and was mixed gently. The cells were then incubated on ice for 30 minutes. Cells were heat shocked for 30 seconds in a water bath set to 42 °C. Each sample was incubated on ice for 2 minutes, and 250 µl of room temperature SOC growth medium was added. Cell recovery was conducted for one hour at 37 °C, with tubes shaking horizontally at 225 rpm. Both 50 µl and 100 µl of the recovered cells were plated to ensure colonies would be present at an optimum density.

Cells were plated out using a sterile plastic spreader onto LB agar containing 50 µg/ml kanamycin; this was done under aseptic conditions. Plates were transferred to a 37 °C incubator for overnight growth.

2.8.6 Overnight liquid culture of cells

5 ml of LB containing the required antibiotic was aliquoted aseptically into the required number of sterile 30 ml glass universal tubes.

The LB agar plates were removed from the 37 °C incubator and care was taken to only open these under a Bunsen flame. If blue/white selection was present, blue colonies present on the plate were disregarded, as their colour indicated that the plasmid they contained had not taken up the PCR insert. Single white colonies were selected, and an autoclaved 200 µl pipette tip was lightly touched to the top of the colony; this tip was used to streak out a fresh plate of this colony, and was then ejected into a 12 ml culture tube. Culture tubes were transferred to a 37 °C incubator with 225 rpm shaking, for 12-16 hours.

Following this period of growth, 1.2 ml of culture was removed and transferred to a Nunc screw-top cryo tube (Thermo Fisher Scientific Ltd.), to which 600 µl of 60% glycerol was added. The tube was inverted several times before being frozen in liquid nitrogen. Tubes of glycerol stocks were transferred to the -80 °C freezer for future regrowth. The remainder of the cultures harvested using a refrigerated centrifuge for 15 minutes at 4,000 *g* at 4 °C.

2.8.7 Plasmid purification

The Wizard Plus SV Minipreps DNA Purification System (Promega UK Ltd.) was used to extract plasmid DNA from bacterial cells. Cells from bacterial culture were harvested by centrifugation for 5 minutes at 10,000 *g*. The resultant supernatant was poured off and excess media blotted off. 250 µl of Cell Resuspension Solution was added to the pellet, cells were resuspended before 250 µl of Cell Lysis Solution was added, and mixed with the lysate by inverting the tube four times. The cell suspension was incubated until the solution had cleared, however the reaction was not allowed to proceed beyond 5 minutes. 10 µl of Alkaline Protease Solution was added to each sample, and was mixed by inverting the tube four times. Samples were incubated for 5 minutes at room temperature; the reaction was not allowed to

proceed beyond 5 minutes to minimise nicking of the plasmid. Neutralising Solution (350 µl) was added to stop the reaction, and the solution was mixed by inverting the tube four times. The bacterial lysate was immediately centrifuged at 14,500 *g* for 10 minutes.

The cleared lysate was transferred to a combined spin column and collection tube. The supernatant was centrifuged at 14,500 *g* for one minute, and the flowthrough discarded before addition of 750 µl of Column Wash Solution to the spin column, and a further 1-minute centrifugation at 14,500 *g*. The flowthrough was once again discarded, and the wash was repeated with 250 µl of Column Wash Solution. The sample was centrifuged for 2 minutes at 14,500 *g* and the flowthrough discarded. A further centrifugation step was conducted to remove any remaining Column Wash Solution. The spin column was transferred to a sterile 1.5 ml microcentrifuge tube, and 50 µl of nuclease-free water was pipetted directly onto the column. The sample was centrifuged at 14,500 *g* for one minute, and the resultant plasmid DNA was quantified using the Nanodrop 1000.

2.8.8 Gateway LR Clonase II reaction

The Gateway LR Clonase II (Thermo Fisher Scientific Ltd.) reaction was conducted to transfer a fragment from the entry vector to the destination vector. This was used with the pBRACT 114 overexpression vector (supplied by BRACT); the pBRACT system utilises a unique ‘dual binary’ vector system, in which the pBRACT vector is co-transformed with pSOUP (Smedley and Harwood 2015). This system was also used with the pGADT7-Rec and pGBKT7 vectors (Takara Bio Europe; supplied by Dr. Barry Causier at the University of Leeds) used for expression in yeast.

The reaction contained 100 ng of the entry vector and 150 ng of the destination vector which were mixed in a 1.5 ml microcentrifuge tube, and the volume was made up to 8 µl by the addition of 1X TE buffer (pH 8.0). The LR Clonase II enzyme mix was thawed on ice for approximately 2 minutes and was vortexed briefly before use. Two µl of LR Clonase II enzyme mix was added to the sample, which was then gently vortexed. The sample was incubated at 25 °C for one hour, at which point 1 µl of Proteinase K was added, and samples were incubated at 37 °C for 10 minutes to terminate the enzymatic reaction. One µl of the LR Clonase reaction was transformed into One Shot TOP10 cells, as described above.

2.8.9 Type IIS restriction enzyme cloning

Type IIS restriction enzymes cleave outside of their recognition sequence, which results in the overhangs that are generated not being dictated by the restriction site. As a result, the design of the resultant overhangs is under the control of the user, and this allows fragments to be joined together seamlessly. This investigation involved the use of the Type IIS enzyme BsaI to bring about the creation of a number of binary vectors that were used for transformation of *A. thaliana*. Some aspects of these constructs were user-generated, and some that were obtained from The Sainsbury Laboratory's collection of 'Golden Gate cloning' modules.

The description of user-generated parts continues in section 2.8.9.1. Vectors obtained from The Sainsbury Lab include the coding region of GFP (pICH41531), a NOS terminator (pICH41421) and the binary vector conferring Kanamycin resistance in plants (pICSL86900). When digested, these vectors resulted in known overhangs to which the overhangs of user-generated parts could be designed to match.

2.8.9.1 Primer design for products to be used in Type IIS restriction enzyme cloning

Each user-generated part was first synthesised by PCR, using primers which added BsaI sites to the 5' and 3' end of the PCR product. The overhangs generated when the fragment was digested were designed to be unique, so that when ligated together, they would only base pair with the complementary sequences of the component they were intended to join with. An example of this process is shown in figure 2.1.

Sequences to be used in conjunction with this method of cloning were checked for the presence of BsaI sites using <http://www.restrictionmapper.org/>. Unwanted BsaI sites present in the DNA were removed through PCR, with the sequence being split into two, and one base being altered to remove the Type IIS restriction enzyme recognition site; this process is shown in figure 2.2. When all unwanted BsaI sites had been removed, the fragments were digested and joined in a single reaction.

2.8.9.2 Digestion and ligation reaction

The reaction was conducted according to the instructions given by SynBIO (<http://synbio.tsl.ac.uk/golden-gate-assembly-protocol/>).

In a 0.5 ml tube, 200 ng of acceptor plasmid was combined with 100 ng of each of the modules to be inserted. To this 1.5 µl of T4 Ligase Buffer, 1.5 µl of 10X Bovine Serum Albumen, 200 units of T4 DNA Ligase and 5 units of BsaI were added (all supplied by New England Biolabs Ltd.); this was made up to 20 µl with nuclease-free water. Samples were transferred to a thermocycler and the following cycling protocol was conducted using a T100 Thermal Cycler: 20 seconds at 37 °C, followed by 26 repetitions of 3 minutes at 37 °C and 4 minutes at 16 °C, 5 minutes at 50 °C, 5 minutes at 80 °C and 5 minutes at 16 °C. 5 µl of the reaction was added to 50 µl of DH5α competent cells, and these cells were transformed as previously described, and plasmid was extracted using the Wizard Plus SV Minipreps DNA Purification System, as described previously.

2.8.9.3 Analysis of constructs

To ensure that the components of each plasmid had assembled correctly, primers were designed to verify the joins in each of the plasmids. These were used in conjunction with single read sequencing provided by Eurofins to ensure constructs were correct before transformation of *Agrobacterium*.

2.8.9.4 Transformation of *Agrobacterium tumefaciens*

The *A. tumefaciens* strain GV3101 conferring resistance to rifampicin and gentamycin was used in conjunction with all plasmids constructed using the method described in 2.8.9.2. The strain was obtained from Ali Pendle in the lab of Prof. Peter Shaw. The *A. tumefaciens* strain AGL1 conferring resistance to rifampicin was used as the host for the *GLO* overexpression vector (*GLO* cDNA in pBRACT114, as described in Chapter 5), and was obtained from Mr. Mark Smedley.

2.8.9.5 Transformation of *A. tumefaciens* strain GV3101

200 ng of plasmid conferring resistance to kanamycin (50 µg/ml) was introduced to 100 µl of GV3101 *A. tumefaciens*. This was incubated on ice for 10 minutes. Samples were cold-shocked by being frozen in liquid nitrogen for approximately 2 minutes, before being transferred to 37 °C for 5 minutes. SOC growth medium (500 µl) was added, and recovery was carried out at 28 °C for 1 hour with shaking at 225 rpm. 100 µl of the recovered cells were plated directly onto agar plates containing rifampicin (50 µg/ml), gentamycin (50 µg/ml) and kanamycin (50 µg/ml). Plates were grown at 28 °C for 2 days.

2.8.9.6 Transformation of *A. tumefaciens* strain AGL1

pBRAC114 (100 ng), and the pSOUP helper plasmid (100 ng), both conferring antibiotic resistance to kanamycin (50 µg/ml) were added to 50 µl of GV3101 *Agrobacterium*. This was incubated on ice for 10 minutes. Samples were transferred to a cuvette and electroporated using a Gene Pulser (Bio-Rad Laboratories Ltd.), set to 400 Ohms, 25 µFD and 200 MV, after which they were immediately transferred to ice and 100 µl of LB medium was added. The contents of the cuvette were moved to a 5 ml culture tube, and recovery was carried out at room temperature for 6 hours with shaking at 120 rpm. Both 50 µl and 100 µl were plated directly onto agar plates containing rifampicin (50 µg/ml), and kanamycin (50 µg/ml). Plates were grown at 28 °C for 2 days.

2.9 *A. thaliana* transformation

The method employed is based on the widely used protocol developed by Clough and Bent (1998). One week prior to transformation, the inflorescences of *Arabidopsis* plants were cut back to encourage the growth of more secondary inflorescences. Immediately before dipping any siliques that had formed were removed.

A single colony of the appropriate *Agrobacterium* strain carrying the desired recombinant plasmid was picked from an LB plate supplemented with the antibiotics to allow growth of the correct *Agrobacterium* strain containing the desired plasmid. This colony was used to inoculate 10 ml of LB supplemented with the correct antibiotics; the culture was grown overnight at 28 °C with shaking at 225 rpm. 1 ml of this culture was added to 500 ml of liquid LB containing the correct antibiotics.

This culture was grown until the OD600 was between 0.6 and 0.8. This was re-suspended in a 5% sucrose solution containing 0.02% Silwett L-77. The inflorescences of plants were dipped in the solution for 5 minutes, and were then laid horizontally and wrapped in cling film to increase humidity. Plants were left overnight, with care taken to ensure they remained in the dark, to increase the transformation efficiency. This dipping process was repeated seven days after the initial dip with the intention of increasing the likelihood of transgenic seeds. Plants were grown in the transgenic greenhouse at the John Innes Centre and were given usual care until seed collection.

2.9.1 Analysis of transformants

Seeds collected from transgenic plants were selected for by growth on 1X MS + 0.8% agar media containing 100 µg/ml kanamycin. Seeds were sown as described previously, but were grown under 24-hour light conditions. Potential transgenics were genotyped via PCR.

2.10 *In situ* hybridisation

2.10.1 Preparation of plasmids for probe synthesis

The sequence chosen to be used as an *in situ* hybridisation probe was first isolated from cDNA by PCR. Primers were designed to highly specific regions of the genes of interest in the hope that they would be unlikely to cross-hybridise to members of the same gene family. These primers were then used in conjunction with Phusion High-Fidelity DNA Polymerase to amplify cDNA template; the use of a proofreading polymerase ensured the accuracy of the synthesised product. PCR products were A-tailed using GoTaq polymerase, which gave them compatibility with the pGEM-T Easy Vector System, allowing ligation into the vector to occur. Ligations were then transformed into DH5α cells using the standard protocol (section 2.8.3).

2.10.2 Linearisation of plasmids for probe synthesis

Before labelling, plasmid DNA must be linearised with a restriction enzyme that leaves a 5' overhang. All enzymes used were supplied by New England Biolabs Ltd. The most commonly used, were NdeI and NcoI-HF. For both enzymes, 2 µl of enzyme, 5 µl of Cutsmart Buffer and 2 µg of plasmid DNA were added; this was made up to 50 µl with nuclease-free water. The reaction was incubated at 37 °C overnight, as these enzymes have been proven to be free from star activity. The digests were heat-inactivated; 20 minutes at 65 °C for NdeI and 30 minutes at 80 °C for NcoI-HF.

2.10.3 Probe-labelling reaction

The RNA DIG labelling kit with T7/SP6 (Roche Diagnostics Ltd.) was used to synthesise digoxigenin-labelled RNA probes. Throughout the probe-labelling reaction the working environment was kept sterile and RNase-free tubes and pipette tips were used. 1 µg of plasmid was made up to 13 µl with nuclease-free water, and to this 2 µl of 10X NTP Labelling Mixture, 2 µl of 10X Transcription Buffer, 1 µl of Protector RNase Inhibitor, and 2 µl of the required RNA Polymerase were added.

Samples were centrifuged briefly and were transferred to 37 °C for two hours. After this incubation 2 µl of DNase was added to each sample to digest the original target DNA. Samples were incubated at 37 °C for 15 minutes, after which time the reaction was stopped with 2 µl 0.2 M EDTA (pH 8.0).

2.10.4 Carbonate hydrolysis of probes

Before use, large probes of over 400 nucleotides in length were carbonate treated to reduce the size to 75-150 nucleotides. The method for calculating the reaction time for carbonation at 65 °C with the intention of reaching a final length of 150 nucleotides is as follows:

$$\text{Time} = (L_i - L_f) / (0.11\text{kb}^{-1}\text{min}^{-1} \times L_i \times L_f)$$

L_i indicates initial probe length and L_f indicates the desired final length in Kb.

To the final volume that resulted from the probe synthesis reaction in 2.10.3, 25 µl of 2 x CO_3 buffer was added, and the probe was incubated at 65°C for the length of time indicated by the equation, based on its initial size. To this, 2.5 µl 10% acetic

acid was added to stop the reaction. To allow for precipitate of the carbonate treated probe, 10 µl of 4M LiCl, 10 µl of 10 µg/µl tRNA (Sigma-Aldrich Company Ltd.), and 300 µl of 100% ethanol were added to the sample, which was mixed and frozen at -20 °C for a minimum of 15 minutes. Samples were spun at 20,000 g in a centrifuge chilled to 4 °C for 20 minutes. The resultant pellet was washed with 100 µl 80% ethanol, spun for 10 minutes, and resuspended in 100 µl 50% deionised formamide (Sigma-Aldrich Company Ltd., made from 50% RNase-free H₂O) before use.

2.10.5 Dot blot testing probes

The labelling efficiency of the probes was measured using a dot blot technique; colour detection was used as the method of quantifying the labelled probe. Both the pre-labelled RNA supplied in the labelling kit and the newly-synthesised probe were diluted by three factors; 1:10, 1:100, 1:10000. 1 µl of each dilution of all samples were pipetted onto a positively charged nylon membrane contained within a sterile square petri dish. Samples were allowed to dry, and the RNA was UV crosslinked to the membrane. The membrane was moistened with wash solution (100 mM Tris-HCl [pH 7.5]; 150 mM NaCl), and 1% Roche Block solution (100 mM Tris-HCl [pH 7.5], 150 mM NaCl, 1% [w/v] Roche block) was added. The membrane was incubated on a rocking shaker for 30 minutes. After incubation the membrane was briefly washed with wash solution, and wash solution containing a 1:3000 dilution anti-DIG-AP (Roche Diagnostics Ltd.) was added to the petri dish. The membrane was incubated for 30 minutes on a rocking shaker. Two 15-minute incubations with wash solution followed, and one 5-minute wash in substrate buffer (100 mM Tris [pH 9.5], 100 mM NaCl, 50 mM MgCl₂) was conducted to equilibrate the membrane to the correct pH. 25 µl of NBT and 15 µl of BCIP were added per 10 ml of substrate buffer, and the membrane was incubated in darkness on a rocking shaker. The colour development reaction usually took no more than 10 minutes. Quantification of probe concentration via this method is not wholly accurate; however the dot blot does give an indication as to whether the RNA probes have been labelled.

2.10.6 Preparation of tissue fixative

A solution of 4% paraformaldehyde was prepared. An 8% (w/v) solution of paraformaldehyde was made up in water, and the solution was heated to 60 °C on a heated magnetic stirrer. 1 M NaOH was added dropwise until the paraformaldehyde

fully dissolved and the solution was clear. 50 ml of 2X PBS was added to the paraformaldehyde solution, and this was left to cool to room temperature. The pH of the solution was checked using indicator strips, and adjusted to pH 7.5 with dilute H₂SO₄ if necessary.

2.10.7 Fixation of plant tissue

Fresh tissue was collected and immediately transferred to freshly-made 4% paraformaldehyde fix solution. The tissue was vacuum infiltrated; pressure was applied to the point at which samples began to boil, and was then slightly released. This was continued until all samples were sufficiently permeated; when the pressure was removed successfully permeated samples would sink to the bottom of the bottle, whereas unsuccessful samples remained floating. After infiltration the fixation solution was exchanged for fresh fixative, and samples were stored at 4 °C overnight. The fixative solution was then exchanged for 70% ethanol, and samples were transferred to cassettes compatible with the Sakura Tissue-Tek® VIP® (Sakura Finetek UK Ltd.), which was used for the subsequent steps of the process.

2.10.8 Processing and embedding of fixed tissue

The dehydration of the plant tissue was performed using a Sakura Tissue-Tek® VIP®, and was achieved through a number of sequential four-hour, 35 °C wash steps: 70% ethanol 80% ethanol, 90% ethanol and three 100% ethanol steps. These washes were followed by three four-hour, 35 °C xylene washes. Finally, four four-hour, 60 °C paraffin wax steps were conducted. Receptacles of each reagent were kept within the unit, and these did not mix. Eosin Y dye (Sigma-Aldrich Company Ltd.) was added to samples to aid sectioning. Samples were then held in wax at 60 °C until required.

2.10.9 Embedding and sectioning of tissue

Samples were embedded in wax using the Sakura Tissue-Tek TEC® Tissue Embedding Console Sakura (Finetek UK Ltd.), and were stored at 4 °C until use. Samples were sectioned to 8 µM thickness using a Leica RM2255 microtome. Sections were transferred to Poly-L-Lysine coated slides (Thermo Fisher Scientific

Ltd.), and were transferred to a warming block at 42 °C. Several drops of sterile water were added to the sample to relax the wax. After approximately half an hour the sample was blotted dry, and was allowed to adhere to the slide at 42 °C overnight. Samples were then stored at 4 °C until required.

2.10.10 Pre-hybridisation

Selected slides were placed in a metal slide rack, and the following washes were conducted to remove wax from tissue sections and subsequently rehydrate the tissue. Slides were subjected to two 10-minute washes with HistoClear to remove the paraffin wax, and were then put through a series of ethanol washes; two 100% ethanol washes for 2 minutes, followed by one of each of the following: 95%, 90%, 80%, 60% and 30%; all steps were of 2 minutes in duration. Samples were incubated in Proteinase K buffer (100 mM Tris-HCl [pH 7.5], 50 mM EDTA) at 37 °C for 20 minutes. A protein digestion was conducted to allow the probe to more easily access the tissues. 1 µg/ml of Proteinase K (Sigma-Aldrich Company Ltd.) was added to Proteinase K buffer pre-heated to 37 °C immediately before slides were submerged. The reaction was stopped by incubation in 0.2% glycine (v/v) in 1X PBS for 5 minutes. A 2-minute wash with 1X PBS followed, before slides were subjected to further fixation in 3.7% (v/v) formaldehyde (Sigma-Aldrich Company Ltd.) made in 1X PBS for 10 minutes. Slides were washed for 2 minutes in 1X PBS twice. A wash of 100 mM triethanolamine, (pH 8.0) with 0.5% (v/v) acetic anhydride was conducted to remove the positive charges from the slides, reducing background levels of staining. Two 5-minute 1X PBS washes followed, and the ethanol series was reversed, from 30% to 100%. Following washing, slides were allowed to air dry.

2.10.11 Hybridisation

Hybridisation solution was prepared, per slide: 0.1-1 ng probe, 0.3 M NaCl, 2X TE buffer [pH 7.0], 0.5 µg/ml tRNA (Sigma-Aldrich Company Ltd.), 0.1 µg/ml Salmon sperm DNA (Sigma-Aldrich Company Ltd.), 1X Denhardt's reagent, 1% dextran sulphate (Sigma-Aldrich Company Ltd.) and 50% deionised formamide (Sigma-Aldrich Company Ltd). This combination of solution and probe was heated to 80 °C for 2 minutes, before being cooled on ice for a minimum of 5 minutes. The cooled solution was pipetted onto the surface of the sections, and a plastic coverslip (Sigma-Aldrich Company Ltd.) was placed on each slide. Samples were placed in a humid

chamber and transferred to a 50 °C water bath where they remained for a minimum of 12 hours.

2.10.12 Post-hybridisation

Slides were removed from the hybridisation chamber and were soaked in 3X SSPE warmed to 45 °C to remove the plastic coverslips. Samples were then transferred to a metal slide rack, and were subjected to three 30-minute washes of 3X SSPE warmed to 45 °C for 30 minutes. A 20-minute wash of 1X NTE was conducted at 37 °C. RNase A was added to a concentration of 20 µg/ml and a 30-minute incubation was conducted at 37 °C. Two 5-minute washes of 1X NTE were conducted. Three 30-minute washes were conducted at 50 °C, of 1.5X SSPE, 1X SSPE and 0.5X SSPE. A 10-minute wash at room temperature with Wash Buffer 1 (100 mM Tris-HCl [pH 7.5], 150 mM NaCl) with 0.1% (v/v) Tween 20 and was followed by a 30-minute wash of Wash Buffer 1 containing 0.5% Roche Block powder (w/v) (Roche Diagnostics Ltd) with mild agitation. A wash step of Wash Buffer 1 containing 1% BSA (w/v) (Sigma-Aldrich Company Ltd.) was conducted with mild agitation. To this 1:3000 dilution of anti-DIG-AP (Roche Diagnostics Ltd.) was added, and was incubated with mild agitation for two hours. Four 15-minute washes of Washing Buffer 1 and 0.2% (v/v) Tween 20 were conducted with mild agitation. Slides were incubated with Washing Buffer 2 (100 mM Tris-HCl [pH 9.5], 100 mM NaCl, 50 mM MgCl₂) with 0.1% (v/v) Tween 20 for 2 minutes. The solution was then exchanged for 100 ml of Washing Buffer 2 containing 250 µl of NBT and 150 µl of BCIP (Promega UK Ltd.). Slides were left in darkness overnight for development.

Colour development was stopped by an incubation series; H₂O, 30% ethanol, 50% ethanol, 75% ethanol, 100% ethanol. This series was then replicated in the opposite direction. Slides were preserved using the permanent mounting media Entellan (VWR International Ltd.). A glass pipette was used to drop a small volume of mounting media onto the slide; a glass coverslip was placed over the sample and light pressure was applied. Samples were allowed to dry overnight before viewing under a microscope.

2.11 Sequence analysis

Unpublished *Primula* genome sequence data generated by the Gilmartin lab was accessed through available for private access in the TGAC browser (<http://tgac-browser.tgac.ac.uk/primula/>).

The NCBI BLAST tool (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to align both nucleotide and protein sequences. ExPASy (<http://web.expasy.org/translate/>) was used to translate DNA sequence into protein and to determine the open reading frame. Multiple sequence alignments were carried out using the Clustal Omega tool (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

2.12 Software packages

User-generated plasmid maps (see appendix) were created using either Benchling (<http://www.benchling.com>) or SnapGene software from GSL Biotech (<http://www.snapgene.com>).

CHAPTER 3

Using qPCR to monitor the quantitative changes in gene expression seen across the developmental stages of buds in *Primula vulgaris*

3.1 Introduction

The traditional model for the control of heterostyly in *P. vulgaris* is based upon the premise that the different floral morphologies of the pin and thrum flowers are controlled by three different functions defined as *G*, *P* and *A* (Ernst 1936b). The *G* function controls the difference in the height of the style, the *P* controls the difference in pollen size, and the *A* function controls the difference in anther height (Ernst 1936b). Thrums were proposed to be heterozygous at the *S* locus, containing the one dominant and one recessive allele, and displaying the genotype *GPA/gpa*. Pins were proposed to be homozygous recessive, with the genotype *gpa/gpa* (Dowrick 1956).

It has recently been discovered that a gene duplication event, which occurred approximately 52.7 MYA, resulted in duplication of the floral development gene *GLO* only in the thrum genome; this duplicated version of the gene has been named *GLO^T* (Li *et al.* 2016). Due to the high level of sequence similarity between the two genes, it was originally presumed that *GLO^T* was a thrum allele of *GLO* (Li *et al.* 2007), however it has since been verified that they are spatially separated within the genome. Since this discovery, it has been found that *GLO^T* is situated, alongside four other genes, in a 278 kb region that is not present in the pin genome. This is the only hemizygous region present in the *P. vulgaris* genome, and it is possible that the genes present in this cluster may be involved in the phenotypic differences seen

between pin and thrum. A hemizygous region, in which only one copy of these genes was present, would behave, genetically, in the same manner as a heterozygous region.

Since the identification of *GLO^T* as a duplicated gene, not an allele, preliminary experiments were conducted to assess its expression. In studies conducted prior to this work, semi-quantitative polymerase chain reaction (PCR) experiments conducted by Dr. Jinhong Li have found that *GLO^T* is expressed only in thrum flower buds, and specifically, appears to only be present in whorls two and three (data not shown). This would suggest it follows the same expression pattern as the gene from which it is duplicated, *GLO*. When a transposon is present in the coding region of the gene, as is the case in the short homostyle *P. vulgaris*, the anthers are no longer raised (Li *et al.* 2016).

Of the three functions defined as being responsible for the altered floral morphologies seen in heterostyly, the localisation of *GLO^T*, and the effect it has when mutated, would suggest it is most likely to be responsible for the *A* function, raising the anthers during floral development, and as a result warrants further investigation to confirm or deny this. The work presented in this chapter builds on these preliminary findings, with the intention of determining the detailed expression profile of *GLO^T* as a candidate for controlling an aspect of the thrum floral phenotype.

As so little is known about *GLO^T*, analysing both the sequence and the expression patterns of this gene, both alone and in comparison to *GLO*, will yield information as to its importance and function. Considering its diversification from *GLO* may also provide insight into the function of this gene. In this chapter the quantitative expression of *GLO^T* across the developmental stages of the flower bud will be assessed, however for this to be possible appropriate normalisation genes must first be tested and established. Quantitative expression of *GLO* across the developmental stages will also be measured, and the levels of expression of *GLO^T* and *GLO* will be directly compared.

3.2 *GLO^T* in *P. vulgaris*

Previous work performed by Dr. Jinhong Li. had shown that *GLO^T* is only present in the thrum *S* locus region of the genome and expressed only in the flowers. As a result, a PCR reaction was conducted on cDNA taken from both pin and thrum buds of mixed stages, using primers designed to the 5' and 3' UTRs of *GLO^T*, and of *GLO* as it is known that this is expressed in both pin and thrum flower buds (Li *et al.* 2008). As shown in figure 3.1, *GLO^T* is expressed only in thrum flower buds. The products of these reactions were sequenced via single read sequencing to allow sequence comparisons to be carried out.

Both *GLO* and *GLO^T* can be categorised as Type II MADS box proteins; from the N-terminus they contain a MADS box, Intervening, Keratin-like and C-terminal domain (Becker and Theißen 2003). The arrangement of the functional domains in this type of MADS box protein has resulted in their also being known as MIKC transcription factors. Recent work has suggested that the Keratin-like domain can be further divided into three separate domains, Keratin-like 1, 2 and 3 (Yang *et al.* 2003).

A comparison between the protein sequences of *GLO* and *GLO^T* shows a high level of sequence identity (figure 3.2), with the majority of the diversification having occurred in the keratin-like domains. These are so-called as they are proposed to form a coiled-coil structure much like keratin, and hence aid protein-protein interactions, including MADS-box dimerisation (Davies *et al.* 1996). In the *A. thaliana* proteins AP3 and PI it has been shown that the Keratin-like 1 and 2 domains are predominantly responsible for dimerisation between the two proteins (Reichmann *et al.* 1996; Yang *et al.* 2003); there is one amino acid change between *GLO* and *GLO^T* in the Keratin-like 1 domain, and five in the Keratin-like 2 domain, which could suggest that *GLO^T* dimerises to a different partner.

Expression of *GLO^T* in *P. vulgaris* is present in thrums and not in pins, and as such differs from *GLO*, which is expressed in both flower morphs. However, the presence or absence of the duplicated gene in other *Primula* species was not known; if gene expression were present in other species, this would suggest it has been conserved as it is performing an important function. As a result, the expression of the *GLO^T* homologues was investigated in closely related *Primula* species.

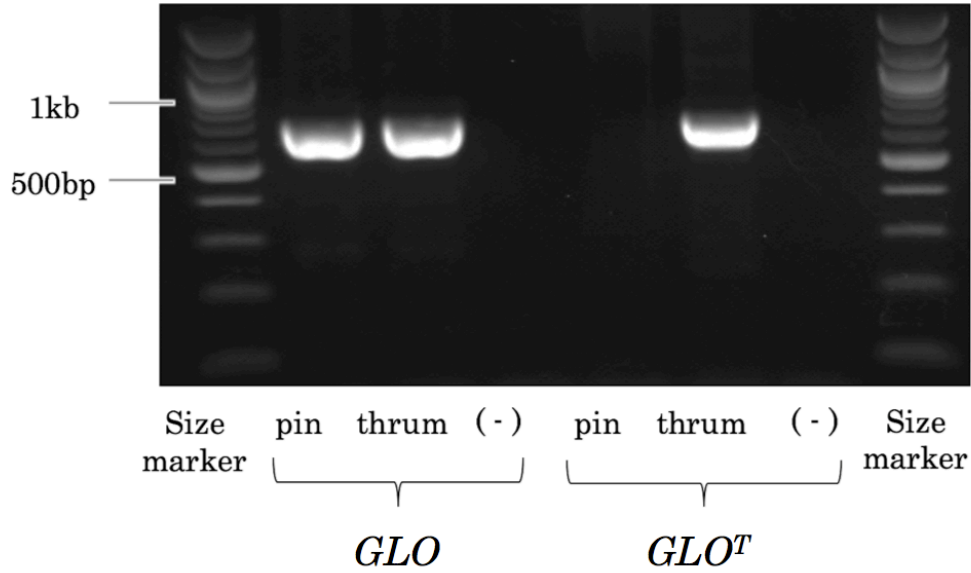


Figure 3.1: Expression of *GLO* and *GLO^T* in the flower buds of *P. vulgaris*. Standard GoTaq PCR reaction using cDNA from mixed stage pin and thrum flower buds alongside no DNA (-) controls. Primers designed to the 5' and 3' UTRs of *GLO* and *GLO^T*. Relevant sizes of bands from the 100 bp ladder are indicated as size markers.

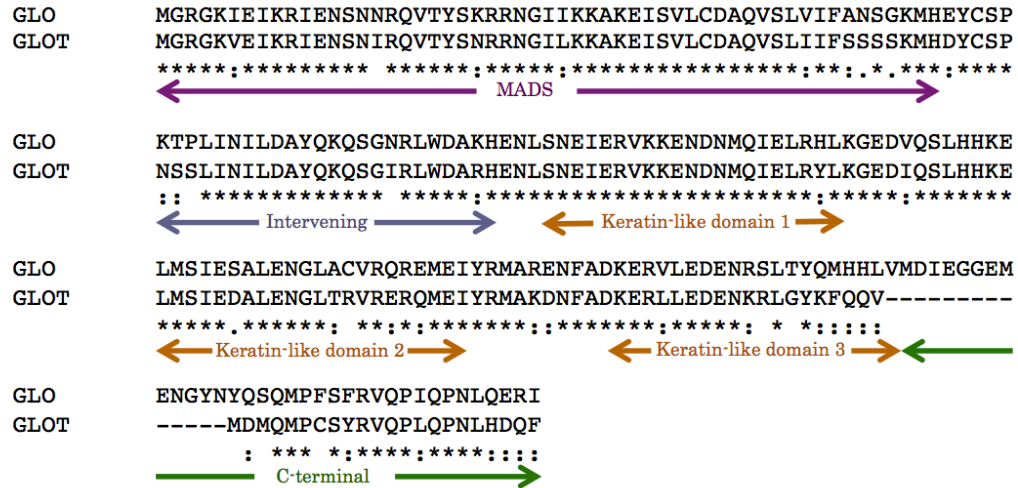


Figure 3.2: Comparison of the amino acid sequences of *GLO* and *GLO^T* in *P. vulgaris*. Domains characteristic of MADS-box proteins are indicated. The * indicates conserved residues, the : indicates conservation between groups of strongly similar properties, the . indicates conservation between groups of weakly similar properties, a blank space indicates no conservation.

3.3 *GLO^T* in *P. veris* and *P. elatior*

At this point in the investigation no genome sequence data were available for *Primula* species other than *P. vulgaris*. *P. veris* and *P. elatior* are the most closely related species to *P. vulgaris* (Mast *et al.* 2006), and as a result it was postulated that the orthologues of *GLO* and *GLO^T* could be conserved to the extent they would be amplified by the primers designed to the *P. vulgaris* sequences.

As shown by figure 3.3, the orthologues were successfully amplified by this method, and were designated *PveGLO* and *PveGLO^T* in *P. veris*, and *PeGLO* and *PeGLO^T* in *P. elatior*; the prefix to the gene corresponds to the species name. At the protein level, all homologues of *GLO* show high levels of sequence identity; there is one amino acid substitution between *GLO* and the other two *Primula* species, the sequences of which are identical. The *GLO^T* orthologues show a slightly higher level of divergence; there are two amino acid substitutions when comparing the *P. vulgaris* and the *P. veris* sequence, two between *P. veris* and *P. elatior*, and four between *P. vulgaris* and *P. elatior*. The fact that in both cases, the *GLO^T* orthologues are showing low levels of divergence would suggest that the function of these genes may be constraining sequence divergence, or that these species are extremely closely related, and have not had time to diverge.

Semi-quantitative PCR experiments conducted by Dr. Jinhong Li have narrowed down the expression of *GLO^T* to the conjoined second and third whorls of the *P. vulgaris* thrum flower buds (data not shown); this localisation of expression is further addressed in Chapter 4. This basic localisation does, however, suggest that if *GLO^T* is involved in floral heteromorphy, it may be responsible for the elevated anthers seen in the thrum form of the flower, performing the *A* function of *GPA*. Nothing is yet known about the quantitative expression of *GLO^T* and how this ties into the elevation of the anthers in the flowers, and as a result these data will be obtained by way of quantitative real-time PCR (qPCR). This quantitative analysis of gene expression will answer the question as to when, temporally, *GLO^T* expression is at its highest, but will not further refine where in the flower bud *GLO^T* is expressed. The spatial localisation of *GLO^T* will be considered in Chapter 4, and together, this information will give valuable insight into the expression dynamics of *GLO^T*, and will give insight into its function.

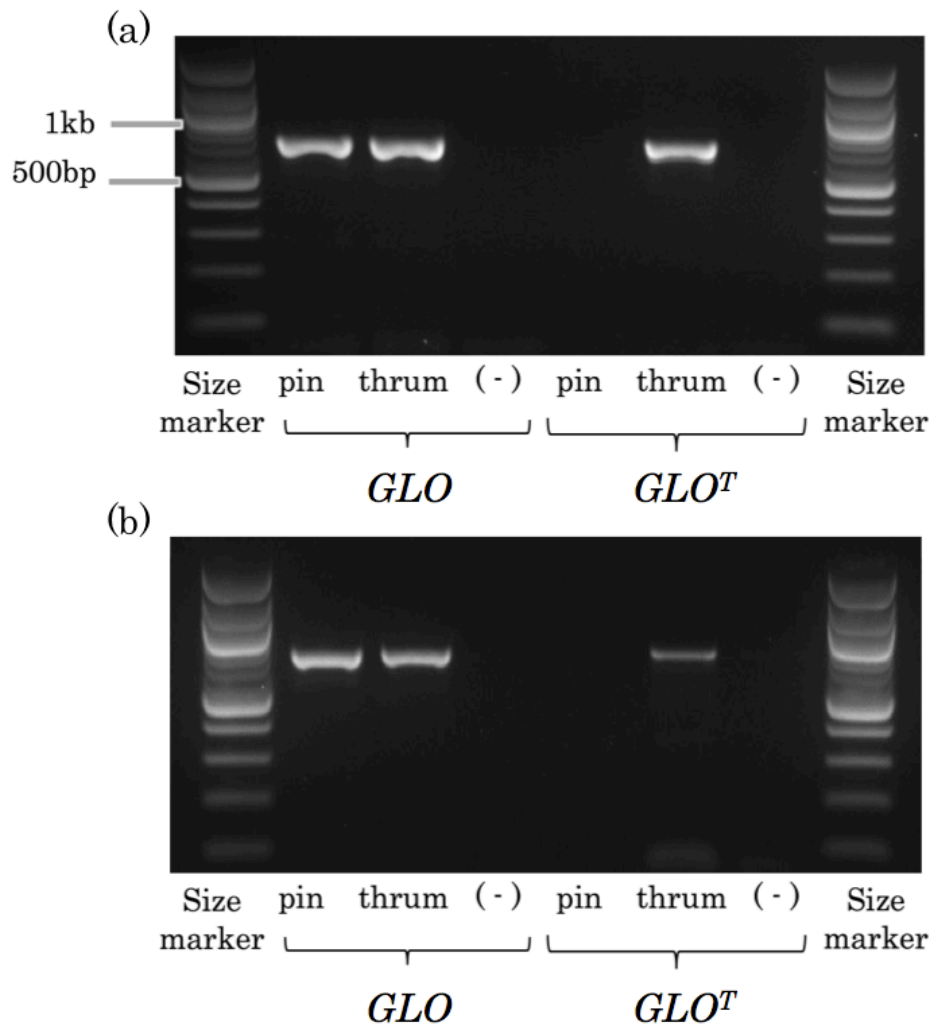


Figure 3.3: Expression of *GLO* and *GLO^T* in *Primula* flower buds. Standard GoTaq PCR reaction using cDNA from mixed stage pin and thrum flower buds. Primers designed to the 5' and 3' UTRs of *GLO* and *GLO^T*. (a) *P. veris* pin and thrum flower bud cDNA, (b) *P. elatior* pin and thrum cDNA. Relevant sizes of bands from the 100 bp ladder are indicated as size markers, and no-DNA PCR controls as (-).

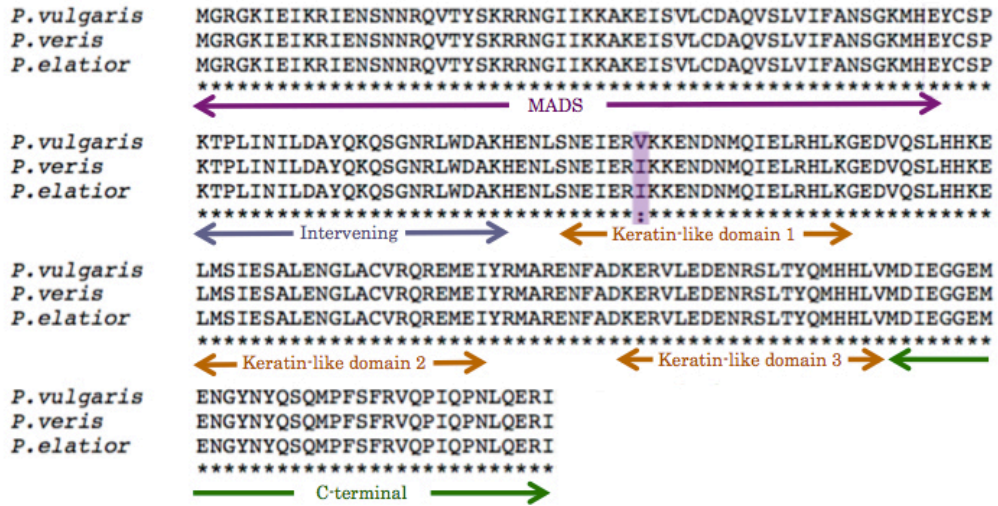


Figure 3.4: Alignment of the amino acid sequence of *GLO* in three *Primula* species. Domains characteristic of MADS box proteins are indicated, and differences in sequence are highlighted in purple. The * indicates conserved residues, the : indicates conservation between groups of strongly similar properties.

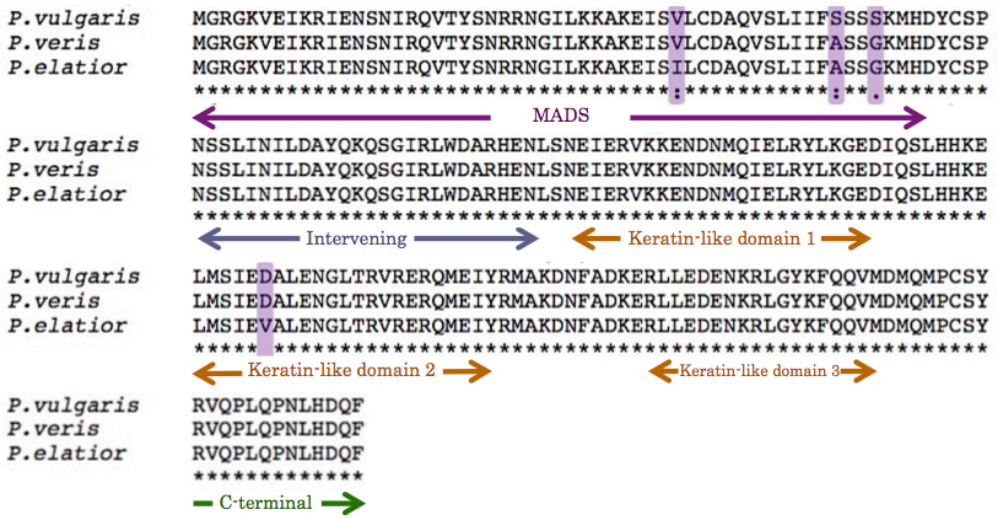


Figure 3.5: Alignment of the amino acid sequence of *GLO^T* in three *Primula* species. Domains characteristic of MADS box proteins are indicated, and differences in sequence are highlighted in purple. The * indicates conserved residues, the : indicates conservation between groups of strongly similar properties, the . indicates conservation between groups of weakly similar properties, a blank space indicates no conservation.

3.4 Selection of *P. vulgaris* flower bud sizes for quantitative real-time PCR (qPCR)

While expression of *GLO^T* had been confirmed in the flower buds of the thrum flowers, the quantification of expression through the development of the flower bud was as yet unknown. Previous work has identified the developmental stages of *P. vulgaris* flower buds; these are defined in Webster and Gilmartin (2006), and will be used to inform this investigation. Bud sizes differ slightly from those stated in the literature, as this investigation is being carried out on wild type *P. vulgaris*, whereas previous experiments were performed using the *P. vulgaris* cultivar ‘Blue Jeans’, which displays a slightly altered flower size in comparison to the wild type. The five different developmental stages being investigated in the wild type *P. vulgaris* are 2 mm, 5 mm, 10 mm, 15 mm and 20 mm flower buds. In the 2 mm buds, the pin stigma has begun to raise above the level of the thrum sigma, however there is no difference in the height of the anthers. This difference in stigma height becomes more prevalent in 5 mm buds, however difference in anther height is still not seen. At the 10 mm stage, the difference in anther height becomes apparent as the corolla tube begins to elongate below the point of anther attachment; the difference between the stigma heights also becoming more prominent. At the 15 mm stage the flower buds of both pin and thrum continue to become more pronounced, being most obvious in the 20 mm flower buds, at which point they are ready to open.

3.5 The selection of appropriate reference genes for qPCR

For data obtained from qPCR experiments to be accurate, they must be normalised to control for differences in cDNA quality and quantity between samples. qPCR is a well-established technique that has been performed on a large number of different species and tissue types, and as a result the experiments conducted in other species are able to inform the choice of reference genes used in these experiments. It has previously been shown that the use of only one reference gene can lead to high levels of error (Vandesompele *et al.* 2002), and as a result, it would be preferable to find a number of reference genes that could be used for normalisation.

At the point at which this investigation commenced, there were no reference genes available for *P. vulgaris*, and as a result, data were collected regarding genes which had been used for the purpose of normalisation in other species. A gene previously identified by Dr. Jinhong Li as showing stable expression across the developmental stages under semi-quantitative conditions was included in the normalisation tests.

This gene was named *P. vulgaris* *ACTIN* (*ACT*), and showed the highest sequence identity to the *A. thaliana* gene *ACTIN11* (*ACT11*), a gene which has been used as a reference gene in both *A. thaliana* and in other species (Mallona *et al.* 2010). A search of the literature revealed a number of gene candidates that showed high levels of stability across various tissue types; one study in particular, Czechowski *et al.* 2005, proved to be a useful resource in this respect. A number of the genes verified by this work to show stable expression in *A. thaliana* were used to obtain the most closely related gene in the *P. vulgaris* genome by way of their DNA or protein sequence, depending on their level of sequence identity; this is discussed in 3.5.1. The initial list of candidates is detailed in table 3.1.

Table 3.1: Initial list of candidates for gene normalisation. Table indicates the name of the *A. thaliana* genes, a description of their functions, and their accession numbers. *denotes the orthologue has already been found in *P. vulgaris*.

Gene name	Description	Accession no.
<i>PEROXIN 4 (PEX4)</i>	Ubiquitin-conjugating enzyme	AT5G25760
<i>ELONGATION FACTOR 1 ALPHA 4 (ELF1α-4)</i>	Calmodulin-binding protein	AT5G60390
<i>PROTEIN PHOSPHATASE 2A-2 (PP2A-2)</i>	Catalytic subunit of protein phosphatase 2A	AT1G10430
<i>MONENSIN SENSITIVITY 1 (MON1)</i>	SAND-family protein	AT2G28390
<i>F-BOX KELCH REPEAT PROTEIN</i>	Galactose oxidase/kelch repeat superfamily	AT5G15710
<i>TUBULIN ALPHA 5 (TUA5)</i>	Isoform of alpha tubulin	AT5G19780
<i>ACTIN 11 (ACT11)*</i>	Actin expressed predominantly in reproductive development	AT3G12110

PEX4, *ELF1 α -4*, *PP2A-2*, *MON1* and *AT5G15710* (details listed in table 3.2) have all been proven to exhibit stable expression across various tissues types in *A. thaliana* (Czechowski *et al.* 2005). Orthologues of these genes have been found suitable for normalisation in various species; orthologues of *ELF1 α -4* and *AT5G15710* have been shown to be suitable for normalisation in leaf and flower development in *Petunia hybrida* (Mallona *et al.* 2010), *PEX4* and *MON1* in maturing embryos of *Brassica napus* (Chen *et al.* 2010), and *TUA5* in *Glycine max* (Hu *et al.* 2009). While all of these genes have been validated in numerous species, it is worth noting that this does not guarantee stable expression in *P. vulgaris* and all genes must first be tested before use. However, due to their proven stability in other species, these represent good candidates to act as controls for gene expression normalisation by qPCR in *P. vulgaris*.

3.5.1 Identifying the *P. vulgaris* homologues of *A. thaliana* normalisation genes

With the exception of *ACT*, the *P. vulgaris* homologues of the genes identified as potential normalisation candidates in table 3.1 needed to be identified. cDNA sequences for each of the genes were identified from The Arabidopsis Information Resource (TAIR, <https://www.arabidopsis.org>), and these were compared by way of a Nucleotide BLAST alignment with the unpublished *P. vulgaris* genome sequence, accessed via the TGAC browser (<http://tgac-browser.tgac.ac.uk/primula/>). The only gene for which a homologue was successfully identified in *P. vulgaris* was the *ELONGATION FACTOR 1 ALPHA 4 (ELF1 α -4)* gene. The homologue was named *ELF1 α* .

For the remaining candidates the protein sequence was obtained from TAIR, and these were used by Jonathan Cocker to identify the most closely related gene in the *P. vulgaris* genome sequence. This was performed using the Exonerate protein2genome command, and from this tool the cDNA sequence of the most closely related gene was made available. This cDNA sequence was compared by way of a BLASTN alignment to the unpublished *P. vulgaris* genome sequence in order to obtain the intron positions, allowing for design of primer pairs which could not amplify genomic DNA.

Of the genes listed in table 3.1, it was not possible to find a candidate that closely resembled *MONENSIN SENSITIVITY 1 (MON1)* or the *F-BOX KELCH-REPEAT PROTEIN (AT5G15710)*. It is possible that at the time of this investigation

the *P. vulgaris* genome sequence was not contiguous over the regions which contained these gene homologues, or that the homologues in *P. vulgaris* have diverged to such an extent that they would not align with the *A. thaliana* sequence.

3.5.2 Testing the *P. vulgaris* homologues for specificity

Table 3.2 lists the candidate genes for which a potential *P. vulgaris* homologue was found, and to what level it is similar to its *A. thaliana* homologue. For each gene, a number of primer pairs were designed, spanning introns to ensure only cDNA targets were amplified; nucleotide sequences of primers are available in the appendix. The sequence of each primer was checked against the *P. vulgaris* genome sequence to ensure that they would not amplify similar genes, however due to gaps in the genome sequence it was deemed necessary to test all primer pairs under standard GoTaq PCR conditions (section 2.3.2) to determine if all produced single bands. The number of primer pairs that produced only one band is listed in table 3.2.

Primer pairs that produced only one band under standard GoTaq PCR conditions were subsequently subjected to amplification efficiency checks detailed in sections 2.7.2 and 2.7.3. Primer pairs were tested via coefficient of determination (R^2); this uses a serial dilution of cDNA to predict the linearity of the Cq data, and assesses the reliability seen across the technical replicates (Taylor *et al.* 2010). An R^2 value of over 0.980 suggests that a primer pair has desirably high amplification efficiency (Taylor *et al.* 2010). A melt peak analysis in which only one distinct peak is present, and hence a product of only one size is being amplified, was also essential for a primer pair to be considered acceptable. The results obtained from the primer pair *ELF1 α* 2F2R are shown in figure 3.6 as an example; for primer pairs that did amplify efficiently the R^2 values for the remaining primer dilution curves are listed in table 3.3, while the remaining melt peak analyses are included in the appendix, as are the primer sequences.

Table 3.2: Refined list of candidates for gene normalisation. Table indicates the *A. thaliana* gene acronym, the chosen acronym for the *P. vulgaris* homologue, the percentage similarity of the cDNA sequences, number of introns present in the *P. vulgaris* homologue, and number of primer pairs which under standard PCR conditions gave a single product of the correct size.

<i>A. thaliana</i> gene name	<i>P. vulgaris</i> homologue name	% similarity of <i>P. vulgaris</i> cDNA sequence to <i>A. thaliana</i>	No. of introns	No. of primer pairs designed (No. that gave a single product)
<i>PEX4</i>	<i>PEX4</i>	80%	3	2 (2)
<i>ELF1α-4</i>	<i>ELF1α</i>	87%	1	2 (1)
<i>PP2A-2</i>	<i>PP2A</i>	82%	5	4 (4)
<i>TUA5</i>	<i>TUA</i>	81%	4	4 (4)
<i>ACT11</i>	<i>ACT</i>	83%	5	5 (4)

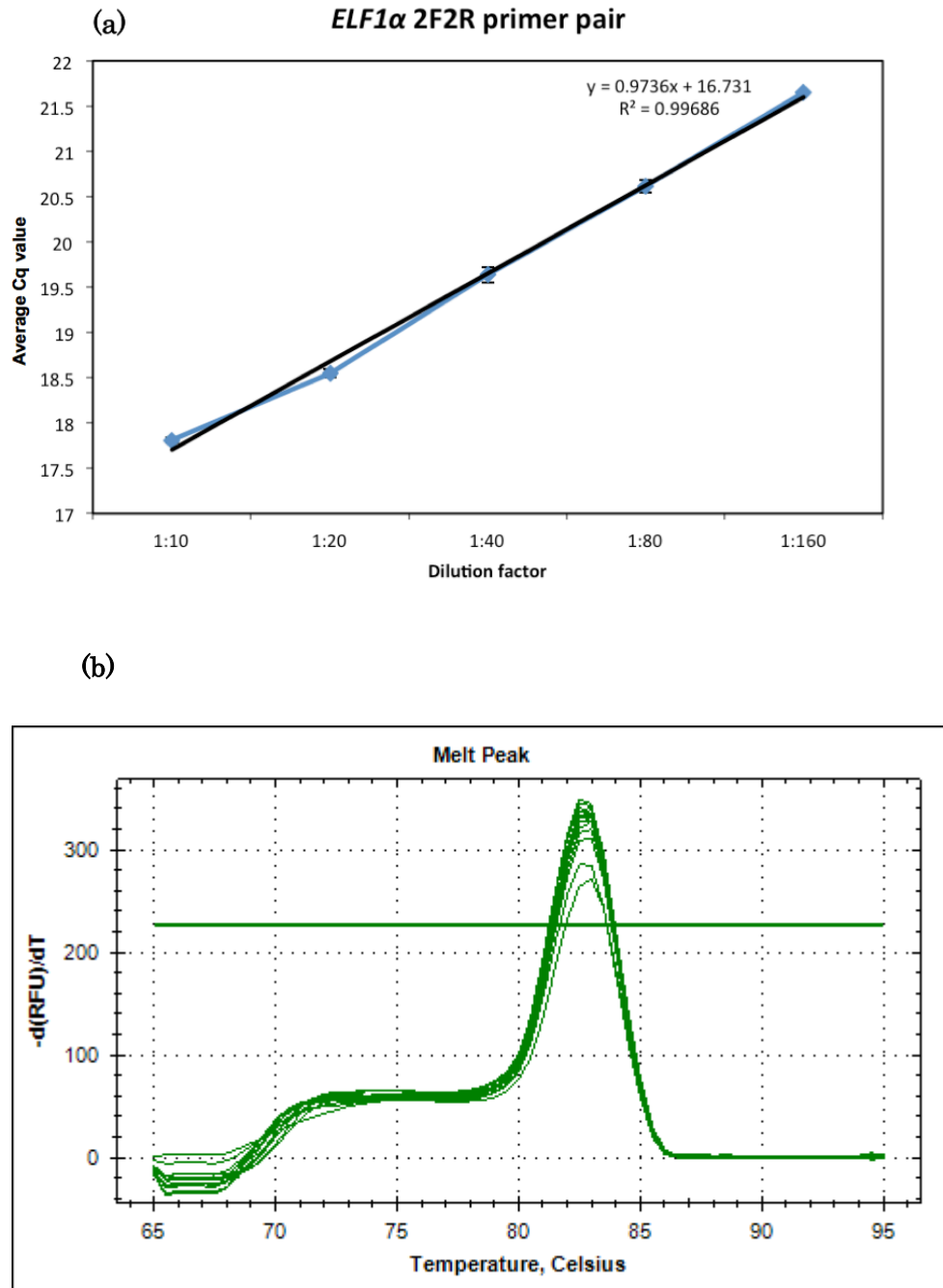


Figure 3.6: Example primer efficiency check and melt peak analysis. (a) Dilution factor plotted against the average Cq value, performed on three technical replicates for each cDNA dilution. Error bars are standard deviation of the mean on these technical replicates. (b) Melt peak analysis conducted at the end of the qPCR run.

Table 3.3: R² values for primer pairs of potential reference genes. The most efficient primer pair is shown in bold.

Gene name	Primer pair name	R ² value	Single melt peak?
<i>PEX4</i>	1F1R	0.992	Yes, but potential primer dimers
	2F2R	Not efficient	Yes
<i>ELF1α</i>	2F2R	0.997	Yes
<i>PP2A</i>	1F1R	0.997	Yes
	2F2R	0.966	Yes
	3F3R	0.974	Yes
<i>TUA</i>	1F1R	0.995	Yes
	2F2R	0.995	Yes
	3F3R	0.997	Yes
	4F4R	0.993	Yes
<i>ACT</i>	1F1R	0.996	Yes
	2F2R	Not efficient	Yes
	3F3R	0.997	Yes
	4F4R	0.981	Yes
	5F5R	0.997	Yes

3.6 The potential for normalisation

For a gene to be a suitable candidate for normalisation, it must be expressed to a highly similar level in all of the tissue samples being investigated. A number of software tools have been developed to statistically verify the stability of potential normalisation genes by way of pairwise testing, varying from the incredibly simple, such as the ΔCt method (Silver *et al.* 2006), to complex pair-wise analyses that are able to select more than one optimal reference gene, such as NormFinder (Andersen *et al.* 2004) and qBase+ (Hellemans *et al.* 2007). All of the available tests perform slightly different analyses to verify the stability of gene expression, and the strategy of using several of these methods to obtain the best combination of reference genes has been widely used (Shivhare and Lata 2016; De Spiegelaere *et al.* 2015; Mallona *et al.* 2010).

Having identified several potential normalisation genes, these were narrowed down to four good candidates for which there are primer pairs that have been shown to efficiently amplify only one product: *ELF1 α* , *PP2A*, *TUA*, and *ACT*. The levels to which these genes are expressed in the flower buds of thrum *P. vulgaris* will be tested and the stability of expression for each gene will be compared against the other potential candidates.

3.6.1 Analysis of raw Cq values

Before detailed testing of gene expression stability was conducted for each of the normalisation candidates, a simple statistical analysis was performed to ensure that all candidates did not show extreme variation when tested on all of the bud sizes required for this experiment. Such variation would automatically exclude them from further testing, and would require that new genes be added to the pool, and go through the tests already performed.

This method has its drawbacks and should not be relied upon as giving a finite ruling on the suitability of a normalisation candidate, as even if cDNA is standardised, this cannot guarantee that all variation between samples will have been removed, and a lack of pair-wise testing leaves these genes un-normalised. It has been previously suggested that raw Cq values with a standard deviation of more than 1.5 are unacceptable for normalisation (De Spiegelaere *et al.* 2015), and this threshold will be applied to these data.

The standard deviation was calculated from all raw Cq values within each biological replicate; these results are shown in table 3.4. *PP2A* and *TUA* display the least variation, with the mean of all three biological replicates being only 0.19 Cq units for *PP2A* and 0.27 for *TUA*. *ELF1 α* shows the next highest level of variation in Cq, and *ACT* shows the highest. The average standard deviation of the Cq values across all three biological replicates is for *ACT*, and the lowest for *PP2A*. All values were under the exclusion threshold of 1.5, and as a result, all potential normalisation genes were subjected to pair-wise testing, to more accurately determine the stability of each gene across the developmental stages.

Table 3.4: Raw Cq analysis of normalisation genes. Standard deviation of the raw Cq values as a measure of gene expression variability between samples. Each biological replicate was first tested independently, before variation across the biological replicates was assessed.

Gene name	Standard deviation of the Cq values			
	Biological replicate 1	Biological replicate 2	Biological replicate 3	Mean of biological replicates
<i>ELF1α</i>	0.37	0.30	0.42	0.36
<i>PP2A</i>	0.32	0.12	0.12	0.19
<i>TUA</i>	0.24	0.28	0.29	0.27
<i>ACT</i>	0.47	0.84	0.77	0.69

3.6.2 ΔC_t method of normalisation gene selection

The ΔC_t method uses the relative expression of pairs of candidate normalisation genes to determine the stability of each gene. This method requires the raw Cq values from all samples, and each gene is normalised against all of the other genes undergoing testing, and from this the stability of each of the candidates was ranked from most to least stable based on how much variation is seen between the samples (Silver *et al.* 2006).

The main advantage of using the ΔC_t method over the raw Cq values is that the candidate genes are being normalised against each other; the differences in cDNA quantity that may exist between the biological samples is controlled for when all genes are tested against each other. Another advantage is that if one gene shows particularly low stability, the standard deviation value for all of the pairs it is included in will be high, allowing it to be identified and excluded from the candidate list. The simplicity of this test, in that it does not require the use of complex statistical tools, allows a list of candidate normalisation genes to be assessed whilst taking into account any potential differences in quantity and quality between sample sub-sets.

The mean standard deviations of each of the pair-wise tests are shown in table 3.5. As was the case when using raw Cq data, *TUA* appears to show the most stable expression; the mean standard deviation of the ΔC_q values for the three pairs it was tested as a part of (*TUA/ELF1 α* , *TUA/PP2A* and *TUA/ACT*) is 0.43 across the three biological replicates, and *PP2A* is once again second with an average of 0.50. *ELF1 α* is third, with 0.62, and *ACT* is ranked as the least stable, with 0.80. A disadvantage of this method is that it does not provide a cut-off point as to what is an acceptable standard deviation for each normalisation gene, and as a result can only be used as a ranking from best to worst, and must be used in conjunction with other tests that give a more definitive ruling on the suitability of a normalisation gene.

The two methods thus far employed both consider the stability, and hence normalising potential, each of the reference genes separately. They do not, however, have the ability to assess the stability of pairs of potential normalisation genes. As a result, a statistical analysis which not only looks at each gene separately, but considers the implications of combining reference genes is the next logical step, and as a result NormFinder was used to facilitate this.

Table 3.5: ΔC_q analysis of normalisation genes. Mean standard deviations of the three pairwise tests conducted for each of the potential normalisation genes when employing the ΔC_t method. Each biological replicate was first tested independently, before variation across the biological replicates was assessed.

Gene name	Mean standard deviation of the ΔC_q values			
	Biological replicate 1	Biological replicate 2	Biological replicate 3	Mean of biological replicates
<i>ELF1α</i>	0.44	0.57	0.62	0.62
<i>PP2A</i>	0.34	0.44	0.51	0.50
<i>TUA</i>	0.29	0.34	0.49	0.43
<i>ACT</i>	0.53	0.72	0.93	0.80

3.6.3 NormFinder

Using the mathematical modelling of gene expression, NormFinder is able to give each normalisation candidate a 'stability value' (S) across the different samples being tested (in this case 2 mm, 5 mm, 10 mm, 15 mm and 20 mm flower buds) and is able to rank each gene by normalisation ability. Lower stability values suggest more consistent expression across the sample set (Andersen *et al.* 2004).

When groups of samples are compared (in this case, comparing stability of gene expression across biological replicates), NormFinder is able to select the two best reference genes, and is able to provide a stability value for the combination of both genes. A NormFinder analysis was conducted on each of the biological replicates alone, to assess not only the stability across the different samples within a biological replicate, but to determine whether the same potential normalisation genes show the lowest stability value in each biologically different sample set; the results of these analyses are shown in table 3.6.

The same pattern of stability was seen in all three biological replicates; *TUA* showed the most stable expression (and was tied with *PP2A* in one sample), followed by *PP2A*, then *ELF1 α* and *ACT*. The stability values for both *TUA* and *PP2A* were consistent between replicates, and were themselves very similar; the range of stability values for both *TUA* and *PP2A* was $0.05 < S < 0.08$. Values for *ELF1 α* were worse than those for *TUA* and *PP2A* in all samples, but were themselves not unacceptable. When all biological replicates were combined, *TUA* and *PP2A* were both considerably more stable than the other two candidates, and the stability value calculated by NormFinder for the combination of *TUA* and *PP2A* was 0.012.

One negative aspect of NormFinder is that does not provide a distinct cut-off as to what is an acceptable value for the stability of a normalisation gene. As a result, a method in which the stability of candidate normalisation gene expression was compared to a pre-set threshold was employed.

Table 3.6: NormFinder analysis of normalisation genes. NormFinder gene expression stability values for the candidate normalisation genes. Stability values for each of the biological replicates when tested alone, and when combined.

Gene name	NormFinder Stability value (S)			
	Biological replicate 1	Biological replicate 2	Biological replicate 3	All biological replicates
<i>ELF1α</i>	0.018	0.028	0.023	0.024
<i>PP2A</i>	0.008	0.007	0.008	0.017
<i>TUA</i>	0.007	0.007	0.005	0.014
<i>ACT</i>	0.033	0.045	0.050	0.035

3.6.4 Biogazelle qbase+

qbase+ is a software package designed to facilitate the analysis of qPCR data. It combines the frequently used statistical analysis of GeNorm (Vandesompele *et al.* 2002) with a number of new tools introduced into the first version of the qBase software (Hellemans *et al.* 2007). The qbase+ software uses an improved version of the original geNorm algorithm to provide a stability value (M) for each of the candidate normalisation genes being tested, and a suggested optimal number and combination of normalisation genes. By this measure, genes that have an M value of <0.5 are considered 'highly stable'.

All candidate normalisation genes were tested on each of the biological replicates, and the geNorm stability value (M) assigned by qbase+ was compared between the four normalisation genes; the results of this are shown in figure 3.7, and the values are listed in table 3.7. In all biological replicates qbase+ determined that *ACT* was the least stable of the candidate genes and can be excluded, however there was no consensus as to which of the candidates was the best, or what the best pair of genes was. In biological replicates 1 and 3 *ELF1 α* and *TUA* were considered the best pair, in biological replicate 2 it was *PP2A* and *TUA*. In figure 3.7 it is clear that the stability values of *ELF1 α* , *PP2A* and *TUA* do not vary substantially; in table 3.7 the exact values further clarify this. In both biological replicates 1 and 3 *TUA* is ranked as more stable than *PP2A*, however it is worth noting that the difference in stability values is 0.002 in biological replicate 1, and 0.005 in biological replicate 3.

When comparing each gene across biological replicates, the stability values of *PP2A* are the most consistent, followed by *TUA*, whereas *ELF1 α* shows the highest variation of the three genes.

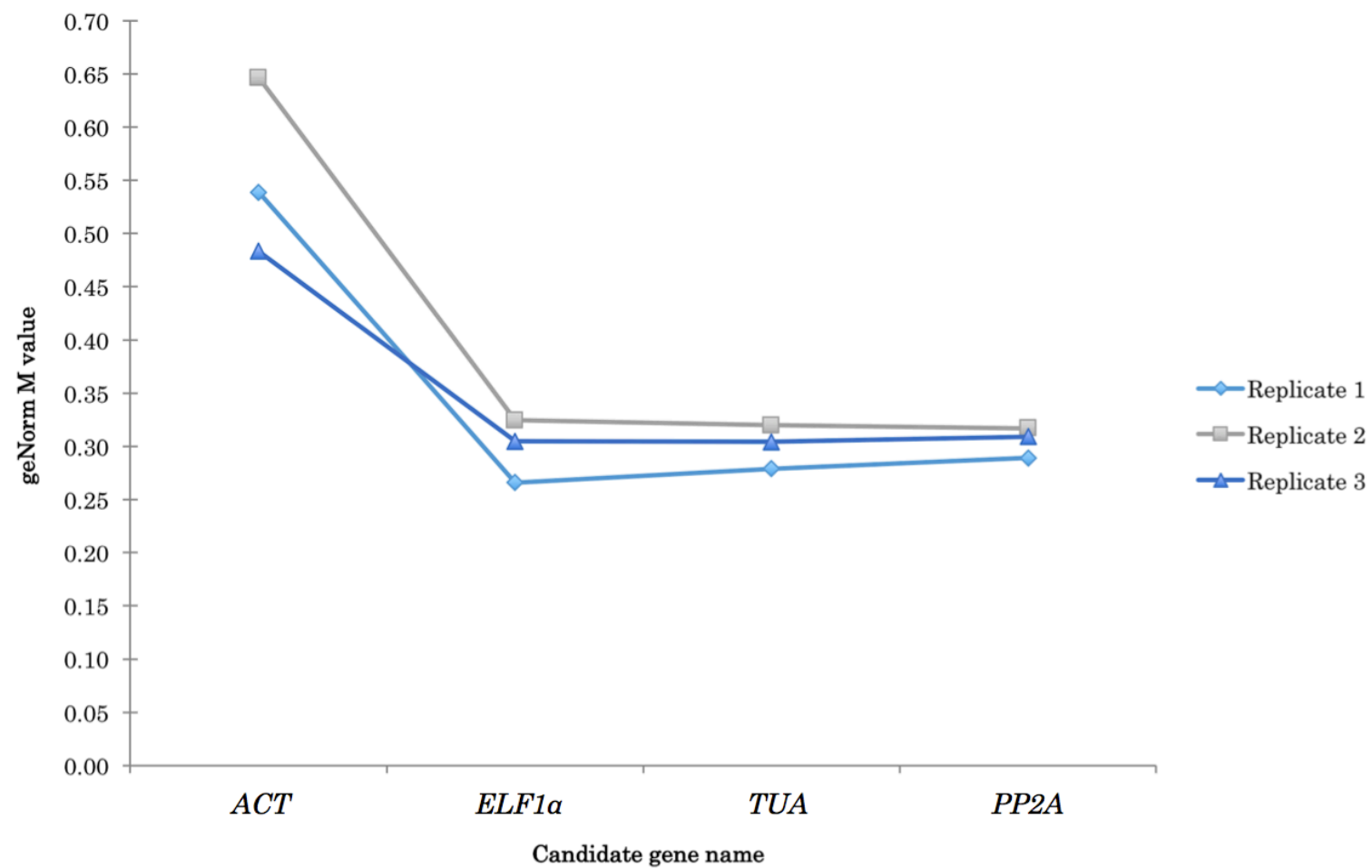


Figure 3.7: Comparison of geNorm M values. geNorm M values of the candidate normalisation genes in the three biological replicates.

Table 3.7: geNorm analysis of normalisation genes. geNorm M values for the candidate normalisation genes in the three biological replicates.

Gene name	geNorm M value		
	Biological replicate 1	Biological replicate 2	Biological replicate 3
<i>ELF1α</i>	0.266	0.325	0.305
<i>PP2A</i>	0.289	0.317	0.309
<i>TUA</i>	0.297	0.320	0.304
<i>ACT</i>	0.539	0.646	0.484

3.6.5 Conclusions that can be drawn from all methods

All methods of assessing the normalising capability of the four candidate genes, *ELF1 α* , *PP2A*, *TUA*, and *ACT*, have produced results, which vary slightly, however the main conclusions that can be drawn from all methods are the same. The first is that *ACT* shows consistently less stable expression across the samples when compared to the other three genes, and as a result should be discarded immediately as a candidate for normalising between the different flower bud stages of thrum *P. vulgaris*. The second is that all three of the remaining genes, *ELF1 α* , *PP2A*, and *TUA*, are suitable for use as normalisation genes. The combination in which they should be used, however, is not immediately clear. Both the Δ Ct method and NormFinder suggested that *PP2A* and *TUA* were better normalisation candidates than *ELF1 α* , and this was consistent across the three biological replicates. This was not the case for qbase+, the results of which were consistent between two of the biological replicates, suggesting that *ELF1 α* and *TUA* are the best pair, whereas the final replicate suggested *PP2A* and *TUA*. The stability values of all three genes are extremely similar in all biological replicates, and all are ranked as very stable.

Thus far, the investigations into the stability of the candidate normalisation genes have shown that *ELF1 α* , *PP2A*, and *TUA* all show high levels of stability, however the intention of this study was to identify the best combination of genes. Due to the disagreement between the data on this subject, a further test of normalisation will be carried out. Using expression data of *GLO^T* across the three biological replicates, samples will first be normalised using the geometric mean of *PP2A* and *TUA*, as suggested by the initial results, and then with all three genes, as the results from geNorm were not conclusive across the three replicates. If the normalisation results differ dramatically, further testing may be needed, or further genes may need to be included in the pool.

3.7 Testing the candidate normalisation genes

Using the normalisation gene combinations discussed above, the expression dynamics of *GLO^T* were examined across the developmental stages: 2 mm, 5 mm, 10 mm, 15 mm and 20 mm thrum flower buds. The normalising ability of the two combinations of genes will first be analysed, and from this the changes in expression of *GLO^T* across the developmental stages will be assessed.

It was important to ensure that the primers designed to amplify the *GLO^T* sequence did not also amplify the extremely similar *GLO*; as a result, these were tested under standard PCR conditions to ensure that no product was seen in the pin cDNA flower bud sample. Once this had been ascertained the primers designed for *GLO^T* were subjected to the tests described above; the R^2 value across a dilution curve was 0.998, and the melt peak is available in the appendix.

Three biological replicates were taken to control for the effects that environmental variables could have on gene expression, and relative levels of gene transcript were normalised against the geometric mean of the validated normalisation genes. All three biological replicates were first treated separately to compare the normalising ability of the different normalisation gene combinations on each of the biological replicates; the results of this are shown in figures 3.8 and 3.9. The variation across the biological replicates was then assessed, and the combinations of normalisation genes were compared with each to ascertain whether the pattern observed was the same when either two or three normalisation genes were used.

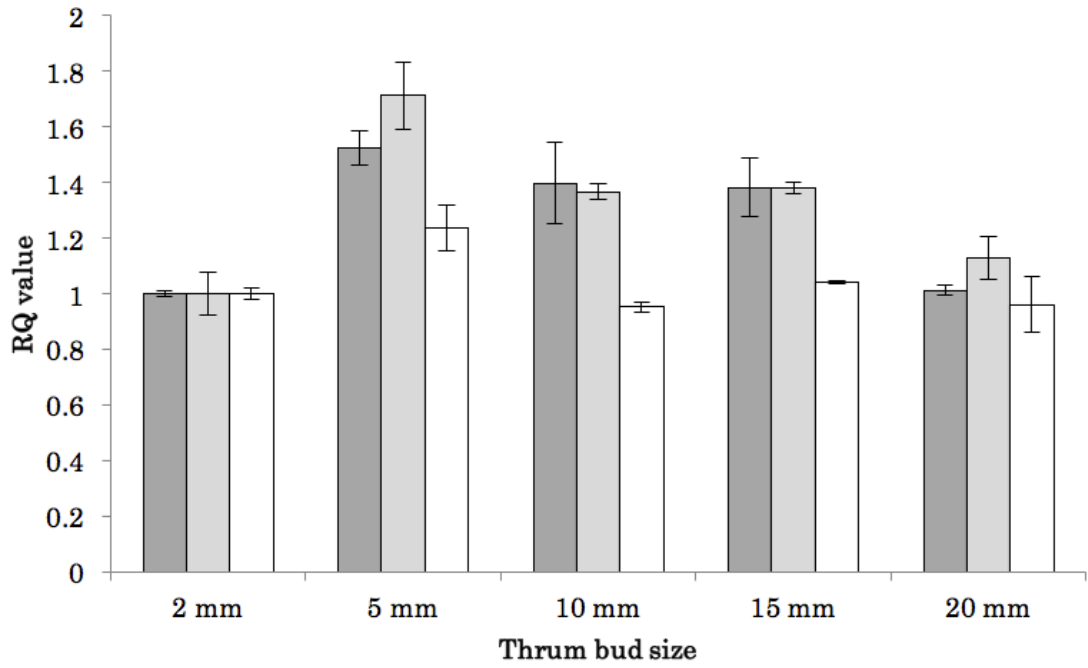


Figure 3.8: GLO^T expression normalised against $PP2A$ and TUA . Expression of GLO^T across the developmental stages of *P. vulgaris* thrum buds, normalised against $PP2A$ and TUA . Data from biological replicate 1 are presented in dark grey, biological replicate 2 in light grey, and biological replicate 3 in white.

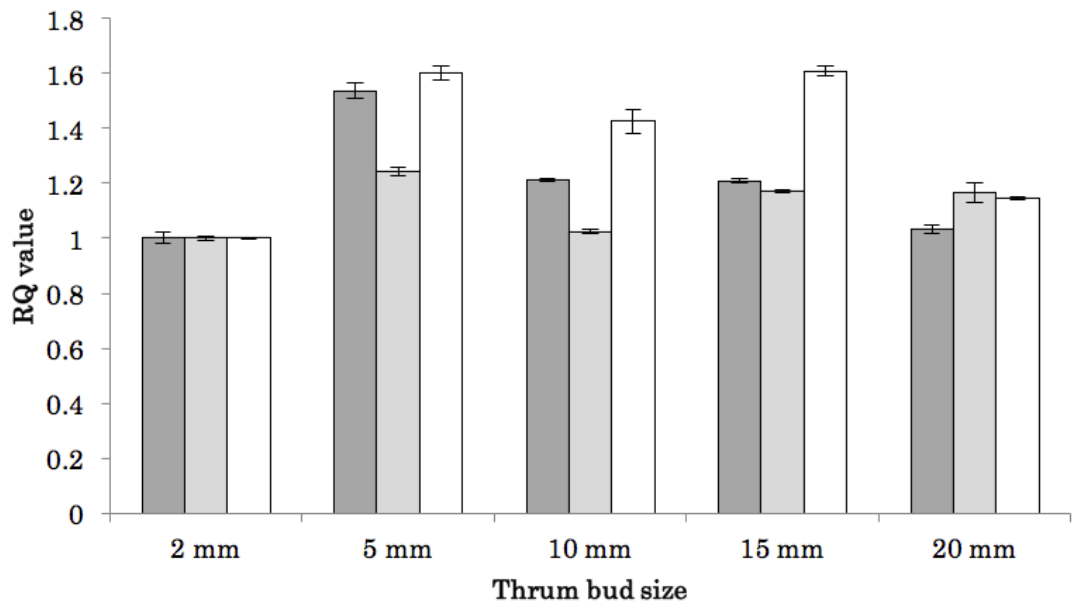


Figure 3.9: GLO^T expression normalised against $ELF1\alpha$, $PP2A$ and TUA . Expression of GLO^T across the developmental stages of *P. vulgaris* thrum buds, normalised against $ELF1\alpha$, $PP2A$ and TUA . Data from biological replicate 1 are presented in dark grey, biological replicate 2 in light grey, and biological replicate 3 in white.

As seen in figures 3.8 and 3.9, regardless of which reference gene combination is used, the difference in expression between the developmental stages is not large in any of the biological replicates; when either two or three reference genes are used, expression of *GLO^T* does not vary by more than 1.7x between any of the samples. When using *PP2A* and *TUA* to normalise rather than the combination of *ELF1 α* , *PP2A* and *TUA*, the differences in expression seen between the developmental stages appears to be more consistent when viewed across the biological replicates, whereas when using the three reference genes, there is a peak at 15 mm for replicate three that is not present in figure 3.8.

When the variation across the biological replicates is assessed, the results are shown in figure 3.10 for normalisation against *PP2A* and *TUA*, and figure 3.11 for *ELF1 α* , *PP2A* and *TUA*. Both figures display the same pattern, suggesting that either combination of normalisation genes will result in the same final trend being displayed.

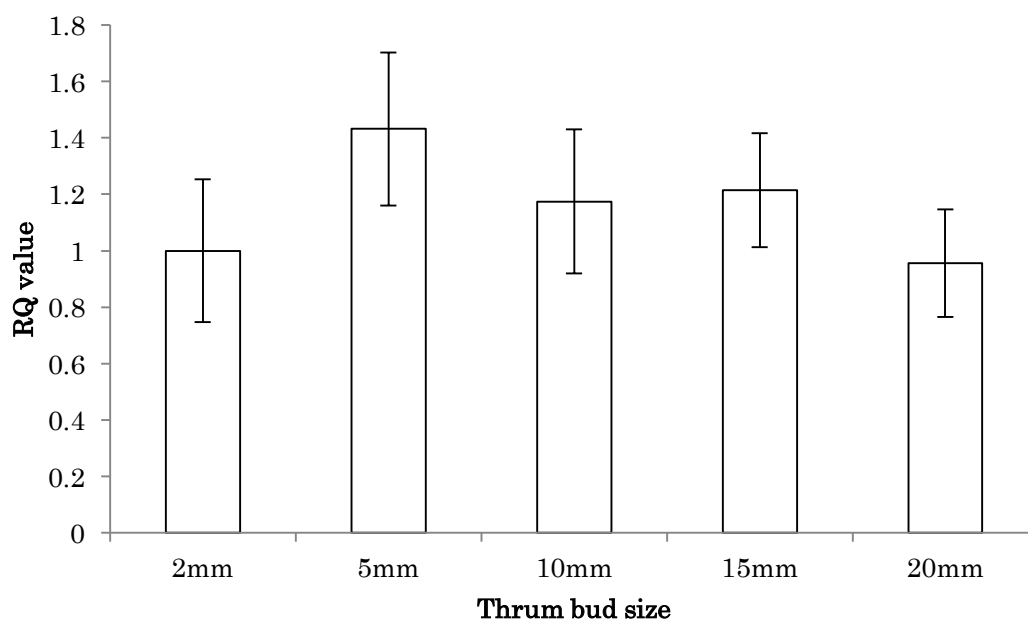


Figure 3.10: Expression of GLO^T across the developmental stages of *P. vulgaris*. Data from each of the biological replicates were collected from separate qPCR runs. Bars show the mean of the three biological replicates normalised against *PP2A* and *TUA*. Error bars show standard deviation.

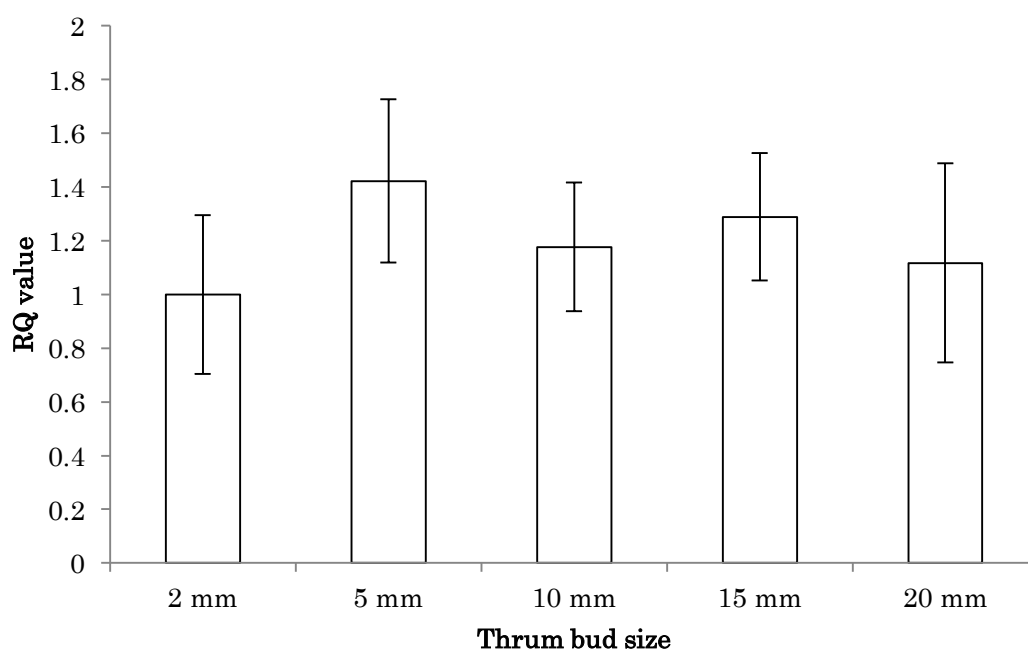


Figure 3.11: Expression of GLO^T across the developmental stages of *P. vulgaris*. Data from each of the biological replicates were collected from separate qPCR runs. Bars show the geometric mean of the three biological replicates normalised against *ELF1α*, *PP2A* and *TUA*. Error bars show standard deviation.

3.8 Assessment of inter-run variation

The differences in *GLO^T* expression between developmental stages of *P. vulgaris* flower buds have been shown to be minimal, however due to high standard deviation of all samples, the slight differences seen between stages are not seen to be statistically significant. It is known that a level of variation exists from one qPCR experiment run to another, and that that can have an impact on results when comparing biological replicates; the $2^{-\Delta\Delta CT}$ statistical analysis method being applied in this investigation does not counteract this (Hellemans *et al.* 2007). These errors can be counteracted by the use of inter-run calibrations, however these are not as accurate as including all samples in the same run (Hellemans *et al.* 2007). As a result, the analysis of *GLO^T* expression was conducted in such a way that all biological replicates were part of the same qPCR run. The intention of this experiment is to monitor the reproducibility of the results, and to see if inter-run variation is influencing these results when not adjusted by inter-run calibrations. Both the combination of *PP2A* and *TUA*, and *ELF1 α* , *PP2A* and *TUA* were tested.

As seen in figure 3.12 and figure 3.13, the general trend of expression seen mirrors that shown in figures 3.10 and 3.11; *GLO^T* expression appears to be stable across the developmental stages, with a slight peak at 5 mm, however this is not statistically significant. When all three biological replicates are run as part of the same experiment, the difference seen between 2 mm and 5 mm buds appears to be less marked. The error bars on both sets of graphs have a high range, and this slight difference is likely to be due to the fact that the changes in expression being measured are low.

Due to the small differences seen in the efficiencies of qPCR reactions between different runs, analyses conducted on data obtained from different runs will contain fluctuations. In order to determine what affect inter-run variation is having on these samples, the standard deviations from both sets of experiments, the former being where each biological replicate was run on a different qPCR run, and the latter being when all biological replicates were included in the same run, were calculated, and are displayed in table 3.8.

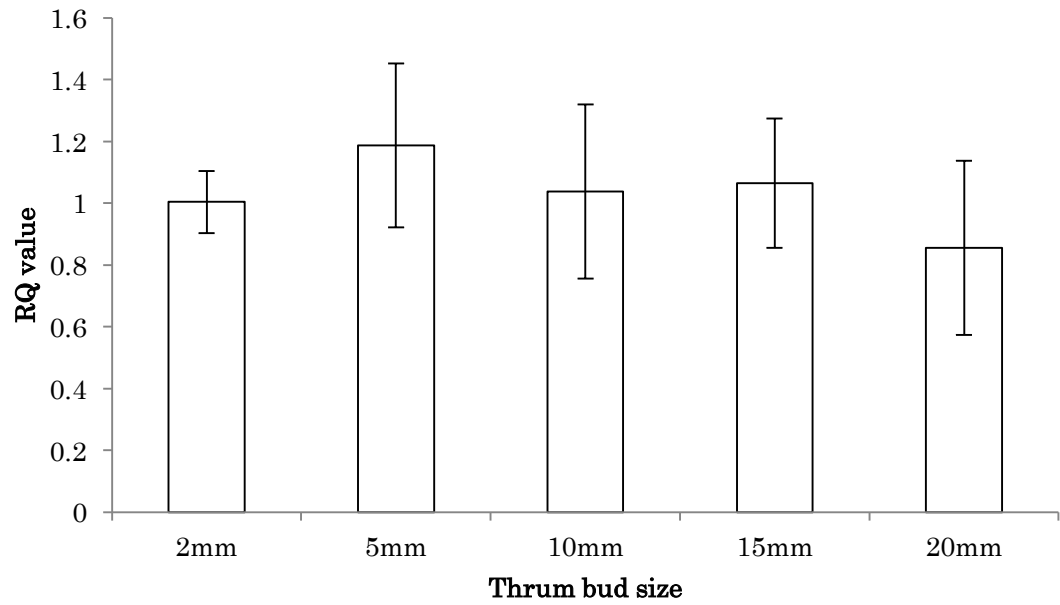


Figure 3.12: Expression of GLO^T across the developmental stages of *P. vulgaris*. Data from each of the biological replicates were collected from the same qPCR run. The mean of three biological replicates was normalised against *PP2A* and *TUA*. Error bars show standard deviation.

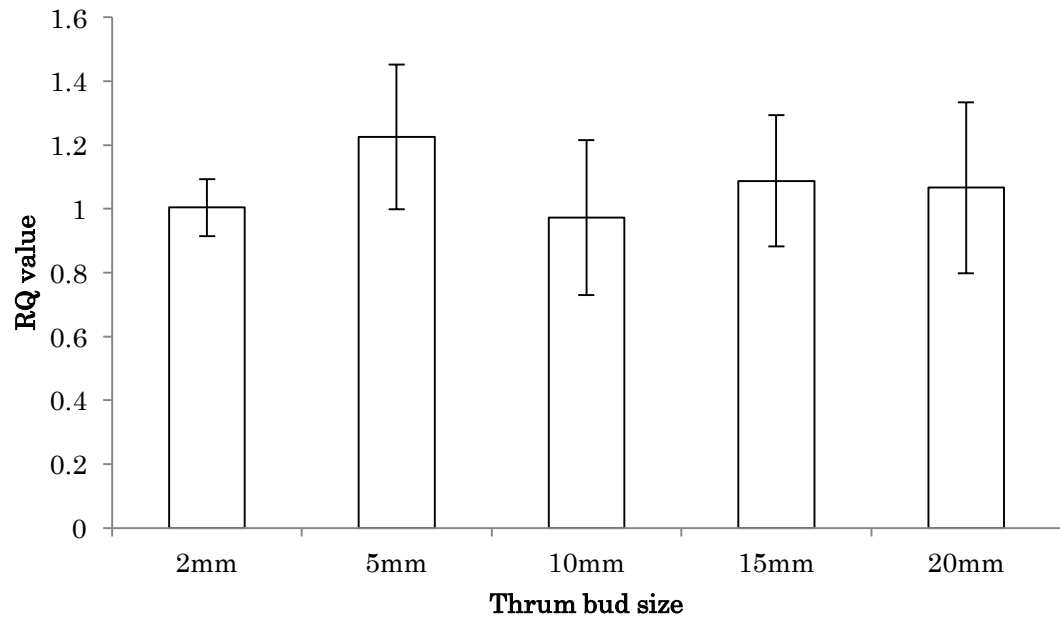


Figure 3.13: Expression of GLO^T across the developmental stages of *P. vulgaris*. Data from each of the biological replicates were collected from the same qPCR run. The mean of three biological replicates was normalised against *ELF1α*, *PP2A* and *TUA*. Error bars show standard deviation.

Table 3.8: Standard deviation of qPCR performed on separate plates combined plates. The average standard deviation was calculated for each of the experimental conditions.

Standard deviation of the mean				
	Separate plates		One plate	
Thrum flower bud size	<i>PP2A</i> and <i>TUA</i>	<i>ELF1α</i> , <i>PP2A</i> and <i>TUA</i>	<i>PP2A</i> and <i>TUA</i>	<i>ELF1α</i> , <i>PP2A</i> and <i>TUA</i>
2 mm	0.253	0.296	0.100	0.0886
5 mm	0.272	0.304	0.264	0.254
10 mm	0.256	0.239	0.282	0.230
15 mm	0.201	0.237	0.210	0.206
20 mm	0.190	0.371	0.282	0.268
Average across all bud sizes	0.235	0.289	0.228	0.209

It has been shown thus far that using two reference genes, specifically the combination of *PP2A* and *TUA*, provides acceptable normalisation, as does the use of three: *ELF1α*, *PP2A* and *TUA*. It is known that the expression of *GLO^T* does not vary substantially across the developmental stages, and that thus far none of the differences between samples have shown statistical significance. A lower standard deviation suggests a higher level of consistency between the biological replicates, and hence better normalisation.

As seen in table 3.8, the average standard deviation for each of the four different methods of normalising the expression of *GLO^T* is lowest when three normalisation genes are used, and when all three biological replicates are performed within the same experiment, removing inter-run variation. The standard deviations for each of the methods of normalisation are not vastly different; the highest is seen when biological replicates are conducted on different plates with three reference genes, and the lowest, as previously mentioned, when three reference genes are used and all biological replicates are included on one plate. As a result, if possible, these are the conditions that provide the optimum normalisation capacity for the genes available.

3.9 The expression patterns of *GLO^T*

The expression of *GLO^T* does not appear to vary greatly across the developmental stages tested. A slight increase in expression of *GLO^T* from 2 mm to 5 mm shows slight variation between 10 mm and 15 mm, and by 20 mm expression levels returns to a level similar to that seen in the 2 mm bud samples. It is important to note that these differences are not deemed statistically significant, suggesting that *GLO^T* is expressed consistently throughout flower development.

The increased anther height seen in thrum flower buds first becomes apparent at between 7 and 11 mm buds in the ‘Blue Jeans’ *P. vulgaris* cultivar (Webster and Gilmartin 2006); in wild type *P. vulgaris* *GLO^T* expression would be expected to be higher in the 10 mm stage than at the 5 mm stage. This would suggest that if *GLO^T* is performing the proposed function of elevating the anthers, it is not direct expression of this gene that is bringing about the change in anther height. As *GLO^T* is a MADS box transcription factor it will be forming either a homo- or a heterodimer to bring about its function. It is possible that the as-yet unknown partner of *GLO^T*

may be undergoing regulation at this stage, whereas *GLO^T* expression remains constant.

3.10 Quantitative expression of *GLO*

While it is known that *GLO* is partially responsible for the determination of floral organ identity in the second and third floral whorls, little is known about its expression in later flower development in *P. vulgaris*. Semi-quantitative PCR conducted before this quantitative analysis showed that expression of *GLO* is detectable in late-stage flower buds (data not shown), however for a detailed analysis of the expression dynamics across the developmental stages, qPCR was performed. Primers designed to the *GLO* sequence underwent the same testing as all others thus described; the primer amplification efficiency curve and melt peak analysis are both shown in the appendix.

As shown in figure 3.14, in which *GLO* expression is measured across the developmental stages in three biological replicates, there appears to be an increase in expression through the developmental stages, with the lowest expression levels being present in 2 mm buds, and the highest in 20 mm. Expression in 20 mm buds is more than 2-fold higher than in 2 mm, and this difference is statistically significant. The expression of *GLO^T* is compared to that of *GLO* in figure 3.15.

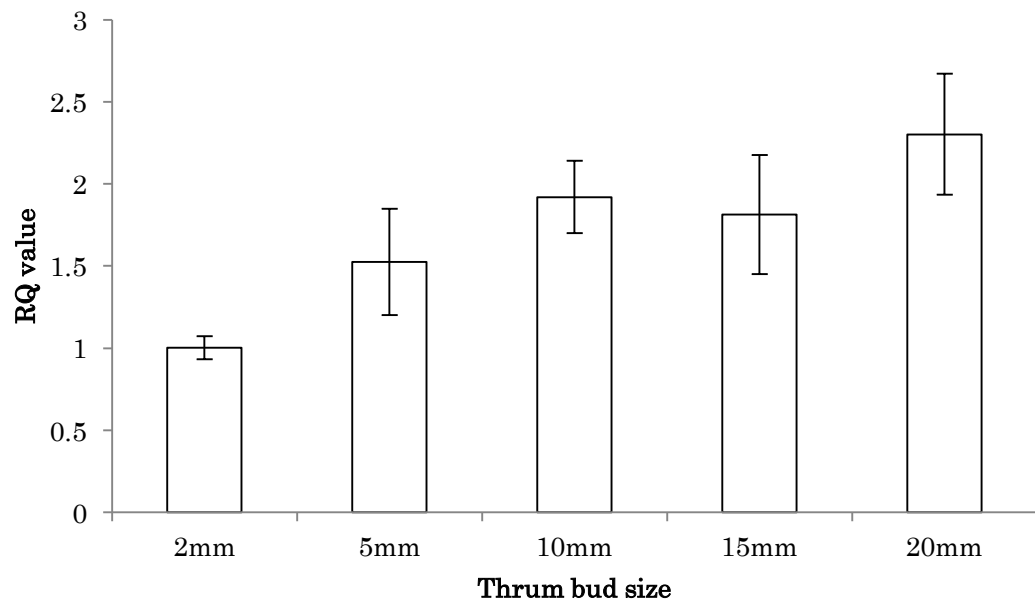


Figure 3.14: Expression of *GLO* across the developmental stages of *P. vulgaris*. The mean of three biological replicates was normalised against *ELF1 α* , *PP2A* and *TUA*. Error bars show standard deviation.

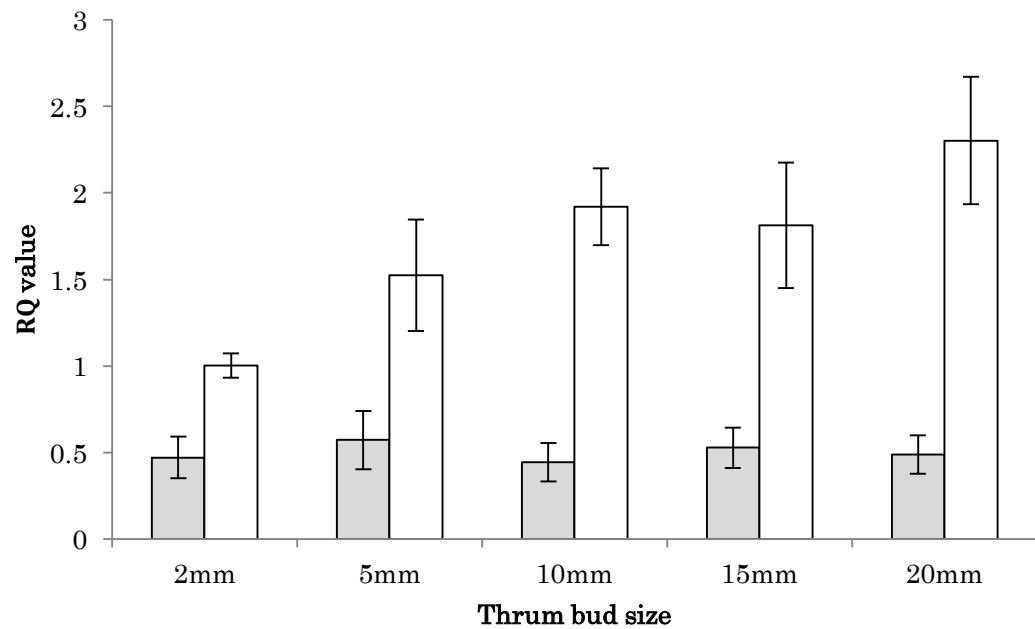


Figure 3.15: Expression of *GLO*^T in comparison to *GLO*. The mean of three biological replicates was normalised against *ELF1 α* , *PP2A* and *TUA*. Grey bars indicate *GLO*^T expression, white bars indicate *GLO* expression. Error bars show standard deviation.

3.11 Comparative expression of *GLO* and *GLO^T*

The expression of *GLO* is significantly higher than *GLO^T* at all of the developmental stages tested; even when *GLO* expression is at its lowest, in 2 mm buds, it is double that of *GLO^T*. As the expression of *GLO* increases through development, the difference in expression between it and *GLO^T* becomes more pronounced; at the 20 mm flower bud stage, expression of *GLO* is more than four times higher than that of *GLO^T*. *GLO* expression has previously been localised to the developing second and third whorls of the floral meristem in *P. vulgaris* (Cook 2002), proving that it is expressed extremely early in development, but not to what level.

3.12 Summary of findings

Multiple tests have determined that, of the available candidates, three suitable normalisation genes have been found to show stable expression across the flower buds of thrum *P. vulgaris*; *ELF1 α* , *PP2A* and *TUA*. These genes will be used in further investigations to assess the expression of the other four genes found at the *S* locus, alongside *GLO^T*.

The expression dynamics of *GLO^T* across the flower bud developmental stages have been elucidated; expression appears to be constant throughout development. This is in contrast to what is seen for *GLO*, which exhibits increased expression across the developmental stages. For *GLO* to be expressed in such a markedly different manner to its paralogue, this may add support to the possibility that the function of *GLO^T* has diverged following the duplication. The localisation of *GLO^T* via RNA *in situ* hybridisation, as seen in Chapter 4, will allow us to assess whether *GLO^T* shows a different spatial expression pattern to *GLO*, in addition to its different temporal expression.

CHAPTER 4

Spatial localisation of *GLO* and *GLO^T* in *Primula vulgaris*

4.1 Introduction

GLO^T is a thrum-specific paralogue of the floral developmental MADS box transcription factor *GLO* in *P. vulgaris*, the expression of which defines the organ identity of the second and third whorls (Li *et al.* 2016). Little is yet known about the expression dynamics of *GLO^T*; in Chapter 3 quantitative analyses were conducted on samples derived from whole thrum flower buds, which indicated that expression does not differ significantly between the developmental stages tested, 2 mm, 5 mm, 10 mm, 15 mm and 20 mm buds, whereas *GLO* expression appears to increase as the flower matures. While this shows that the temporal expression of *GLO* and *GLO^T* differ, these data do not offer any information as to the localisation of these genes within the flower bud.

Early attempts at localising the expression of *GLO* and *GLO^T* involved dissected flower buds in which the four floral whorls had been separated. Semi-quantitative PCR performed by Dr. Jinhong Li showed that, in 15 mm thrum flower buds, expression of *GLO^T* was only detectable in the second and third whorls, and expression of *GLO* has also been confirmed to be present in only the second and third whorls of both pin and thrum flower buds (data not shown). *GLO* has been shown by RNA *in situ* hybridisation to be expressed in the second and third whorls of developing floral meristems (Cook 2002), however the expression of *GLO^T* at this particular stage of flower development needs to be confirmed.

4.1.1 Experimental aims

Localisation of both *GLO* and *GLO^T* expression at the floral meristem stage by way of RNA *in situ* hybridisation will allow for a direct comparison between these two paralogues, and will yield valuable information as to the divergence of function of these two genes. While expression of both *GLO* and *GLO^T* is seen in the conjoined second and third whorls of 15 mm *P. vulgaris* flower buds, *GLO^T* expression is only detectable in thrums due to its absence from the pin genome. *GLO* and *GLO^T* display different levels of expression in the flower buds, as shown in Chapter 3, and it is possible that their expression patterns may also differ during the period in which whorl identity is being defined.

The localisation of the *GLO* and *GLO^T* proteins are also of interest; as of yet nothing is known of their structure or function. Due to the high level of sequence similarity between these *GLO* and *GLO^T*, antibodies raised to full-length proteins were deemed likely to show a lack of specificity. As a result, peptide antibodies were designed for *GLO* and *GLO^T*; this was also done, where possible, for other genes found to be within the *S* locus region (Li *et al.* 2016).

4.2 RNA *in situ* localisation

RNA *in situ* hybridisation uses a complementary, labelled sequence of RNA to localise the expression of that sequence in tissue of interest. Knowing in which region of the tissue a gene is expressed yields useful information as to where it functions. The labelling used throughout this series of experiments is digoxigenin (DIG) labelling, in which molecules of digoxigenin are incorporated into the probe sequence, and are then detected via anti-digoxigenin antibody.

Probes were designed to be as specific as possible to the gene of interest; sequence analysis was conducted to determine the lowest region of sequence similarity. Primers were designed to these regions, a Polymerase Chain Reaction (PCR) experiment was conducted, the PCR fragment was cloned into pGEM-T Easy, and the sequence was validated via single read sequencing to ensure no errors were present. The vector was then linearised and the probes synthesised using the T7 RNA Polymerase promoter. Probes were tested via a dot blot to ensure they were of a high enough concentration to allow visible detection when recognising the sequence of interest and run on a gel before use; the full details of this process are available in section 2.10.

Each probe was synthesised with the sequence of interest in the antisense and the sense orientation. The antisense probe is complementary to the RNA sequence present in the tissue and is able to bind to it, whereas the sense probe is the same orientation as the sequence present in the tissue, and should not hybridise. As a result the sense probe is used as a measure of the background levels of staining seen in each experiment.

4.2.1 RNA *in situ* localisation of *GLO* in floral meristems

Localisation of *GLO* in floral meristems had already been shown (Cook 2002), however at the point at which this work took place, the existence of *GLO^T* was not known. In previous experiments it was not specified whether pin or thrum floral meristems were used in RNA *in situ* experiments, and as a result it is not known whether the *GLO* probe used was cross-hybridising to the *GLO^T* sequence. As a result, it was checked whether the probe used in previous work was specific to the sequence of *GLO*, and the experiment was repeated using same probe sequence as was used in Cook 2002 with the pin and thrum samples marked.

Figure 4.1 shows the aligned coding sequences of *GLO* and *GLO^T*, and the positions of the primers used in previous work to amplify the sequence used to create the *GLO* probe (Cook 2002). The sequence does appear to be significantly different, however, due to the size of the probe, carbonate hydrolysis (section 2.10.4) was performed to ensure that the probe was small enough to penetrate the tissues sufficiently. This hydrolysis may result in shorter sequences which show a high level of homology to both *GLO* and *GLO^T*. If, when tested, the expression appears different between the pin and the thrum samples, this would suggest cross-hybridisation.

As seen in figure 4.2, the expression of *GLO* is shown by the purple staining in whorls two and three of pin and thrum floral meristems, and this is consistent as the floral meristems continue to develop. This expression is proven to be specific, as a control experiment involving sense probes show minimal levels of background staining (also shown in figure 4.2). In order to compare the localisation pattern of *GLO* and *GLO^T* RNA *in situ* hybridisation will be conducted using a *GLO^T*-specific probe.


```

GLO      ATGGGTAGAGGAAAAATAGAGATAAAGAGGATTGAGAACTCAAATAATAGGCAAGTTACT
GLOT      ATGGGGAGAGGAAAGGTAGAGATAAAGAGGATTGAAAACTCGAATATCAGACAAGTGACG
          *****
          *****

GLO      TATTCAAAGAGGAGAAATGGGATCATAAAAAAGGCCAAGGAGATCTCAGTTTATGTGAT
GLOT      TATTCAAACAGGAGAAATGGGATACTGAAAAAGGCCAAGGAGATCTCGGTTTGTGTGAT
          *****

GLO      GCTCAGGTCTCCCTTGTTATTTTTGCCAACTCTGGTAAAATGCATGAATATTGCAGCCCT
GLOT      GCTCAGGTCTCCCTTATTATTTTCTCTAGCTCCAGTAAGATGCATGATTACTGCAGTCCA
          *****

GLO      AAAAATCCGTTGATTAACATCTTGATGCATACCAGAAGCAATCTGGGAACAGGTTGTGG
GLOT      AATTCCTCGTTAATTAACATCTTGATGCATATCAGAAGCAATCTGGGATTAGGTTGTGG
          *****

GLO      GATGCTAAGCATGAGAACCTCAGCAACGAAATAGAAAGGGTCAAGAAAGAGAATGATAAT
GLOT      GATGCTAGACATGAGAACCTTAGCAATGAAATTGAGAGGGTCAAAAAAGAGAATGACAAT
          *****

GLO      ATGCAAATTGAGCTCAGGCACTTGAAAGGAGAAGATGTACAATCTTGCACCACAAGGAG
GLOT      ATGCAGATTGAGCTCAGATACTTGAAGGGAGAAGATATACAATCTTGCACCACAAGGAG
          *****

GLO      CTTATGTCCATTGAATCCGCCCTCGAAAAATGGACTTGCTTGTGTTCCGCAGAGAGAGATG
GLOT      CTCATGTCTATAGAAGATGCACTCGAAAAATGGACTAACTCGTGTTCCGCAGAGACAGATG
          *****

GLO      GAGATTTACAGGATGGCAAGAGAAAATTTTGCTGACAAGGAAAGGGTACTGGAAGATGAA
GLOT      GAGATCTACAGAATGGCAAAAGACAATTTTCGCTGATAAAGAAAGGCTTCTAGAAGATGAG
          *****

GLO      AACAGGAGCCTTACTTACCAAATGCACCACCTGGTGATGGATATAGAGGCGGGGAAATC
GLOT      AACAAGCGCCTTGGCTACAAATTTTCAGC---AAGTGATGGAT-----
          *****

GLO      GAAAATGGATATAATTACCAGTCTCAAATGCCATTTTCCTTCCGGGTGCAACCGATTGAG
GLOT      -----ATGCAGATGCCTTGCTCCTACCGTGACAGCCGCTTCAA
          *****

GLO      CCAAATTTACAGGAGAGGATTTAA
GLOT      CCAAATTTACAGATCAGTTTAA
          *****

```

Figure 4.1: ClustalO alignment of *GLO* and *GLOT* cDNA sequences. The * indicates conserved residues, the : indicates conservation between groups of strongly similar properties, the . indicates conservation between groups of weakly similar properties, a blank space indicates no conservation. Green highlight indicates the start codon; red highlight indicates the stop codon. Yellow highlight indicates the position of the forward primer designed by Cook (2002) and used to create the *GLO*-specific PCR product; blue highlight indicates the position of the reverse primer.

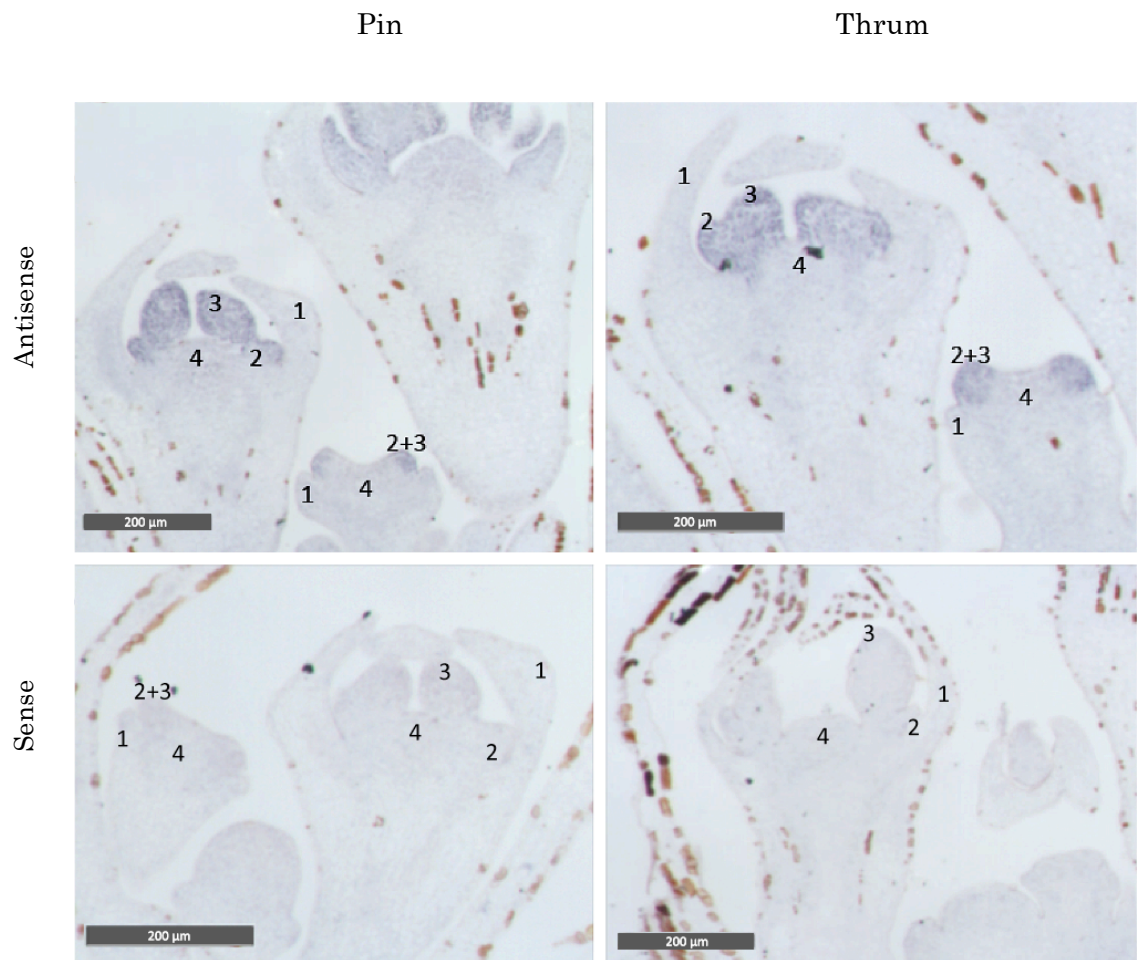


Figure 4.2: RNA in situ localisation of *GLO* in *P. vulgaris* floral meristems. Floral whorls are indicated in the figure by their number (1, 2, 3, and 4). Antisense and sense probes are indicated.

4.2.2 RNA *in situ* localisation of *GLO^T* in floral meristems

The RNA *in situ* localisation of *GLO^T* was first attempted on floral meristem tissue, to allow for a direct comparison between its localisation and that of *GLO*. When attempting to localise expression of *GLO^T*, pin floral meristem tissue was also included. As the sequence of *GLO^T* is only present in the thrum genome, any expression detected in these samples will be due to cross-hybridisation to *GLO*.

Throughout the process of attempting to localise *GLO^T*, cross-hybridisation was found to be a considerable problem, despite numerous optimisation attempts. Multiple different probe concentrations, washing stringencies and fixation strategies were used to reduce this, and the best results were obtained at a high washing stringency with a short probe. The sequences of the primers used in the creation of the *GLO^T* sense and antisense probes are available in the appendix.

As seen in figure 4.3, the expression of *GLO^T* is seen in the floral meristems, but the expression does not appear to be constrained to the same areas as *GLO*. In figure 4.4, there is potentially some cross-hybridisation to other MADS box genes that are expressed in pins, as the antisense pin slides appear to be more stained than the sense despite it being known that *GLO^T* displays no expression in pin tissue, however the antisense samples for the thrum floral meristems are distinctly darker.

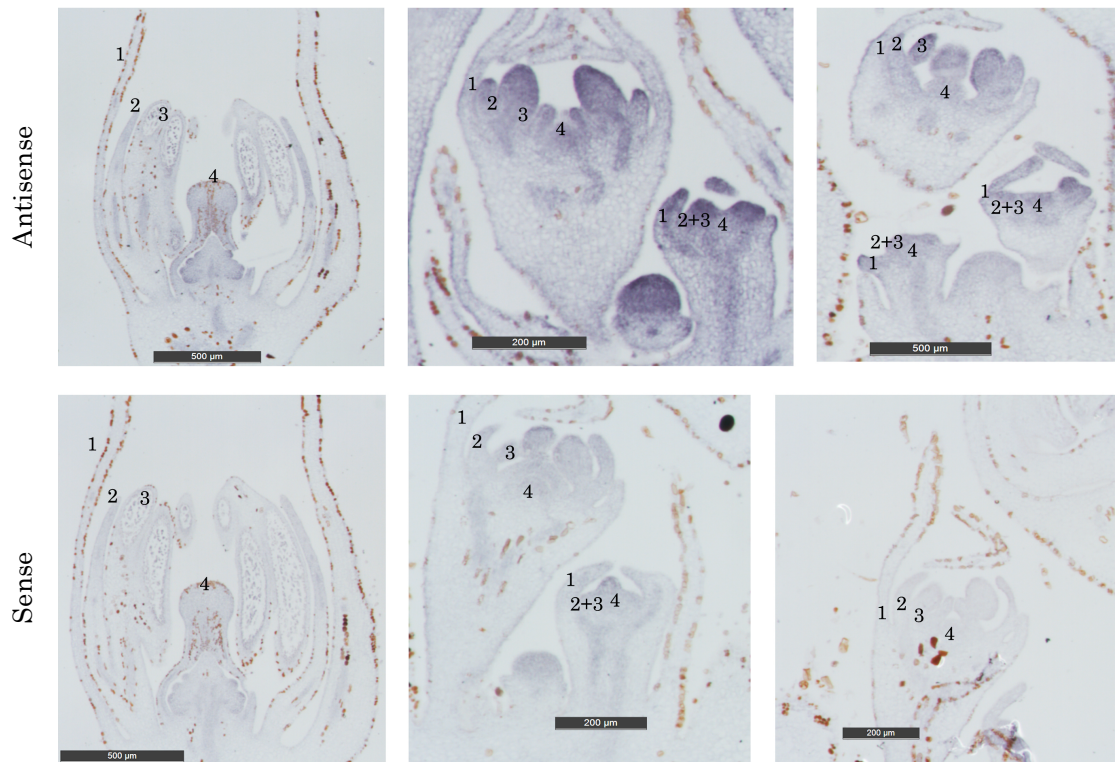


Figure 4.3: RNA *in situ* hybridisation of *GLO^T* probes on thrum floral meristems. Antisense and sense probe use is labelled. Floral whorls are indicated in the figure by their number (1, 1052, 3, and 4).

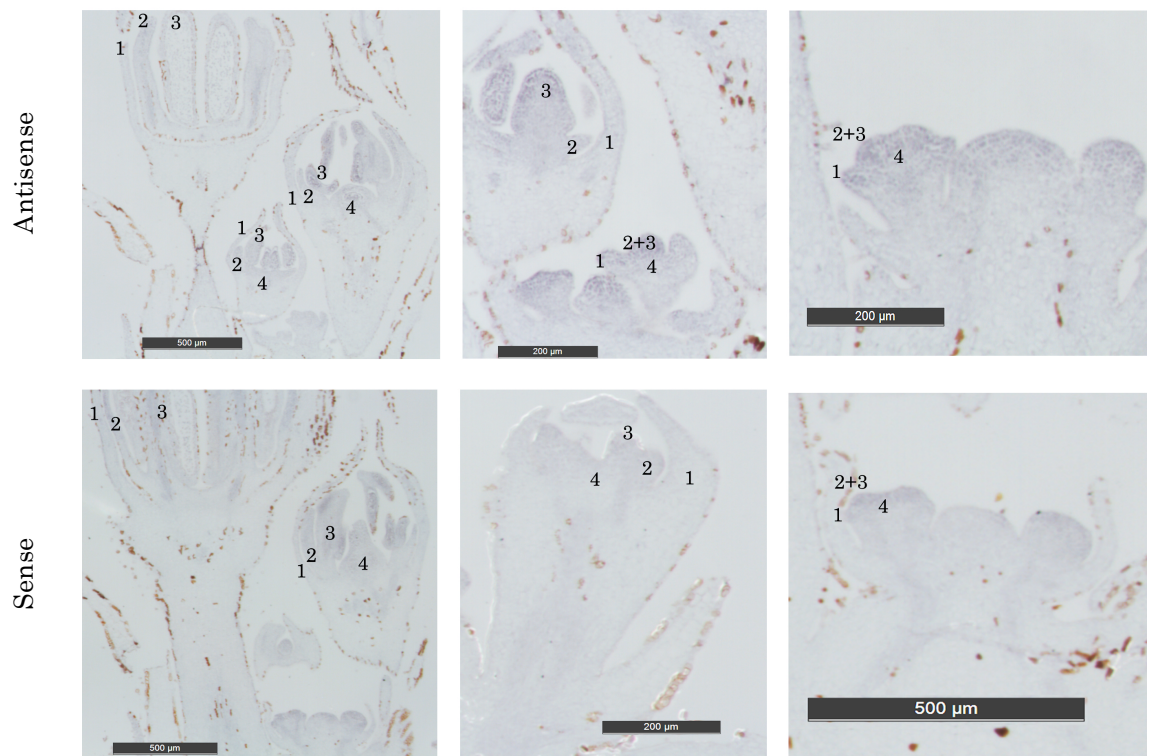


Figure 4.4: RNA *in situ* hybridisation of *GLO^T* probes on pin floral meristems. Antisense and sense probe use is labelled. Floral whorls are indicated in the figure by their number (1, 2, 3, and 4).

4.2.3 RNA *in situ* localisation of *GLO* and *GLO^T* in mature flower bud tissue

Localising expression in expanded tissue, such as that seen in the developing flower buds of *P. vulgaris*, has proven to be challenging, and problems with this technique have not yet been resolved. In large cells with an expanded vacuole, the cell contents become pressed against the cell wall. Numerous attempts were made to localise expression in tissue displaying these cells, using different methods and optimisation steps, however none provided definitive results as to where the localisation of *GLO* and *GLO^T* are found within the flower bud.

Due to the size of the tissue involved, different fixation methods and times were attempted to ensure full penetration. Tissue to be used in RNA *in situ* hybridisation was stained to ensure that, due to expanded cell size, cell content was still present within the cells; Alcian blue was used to stain non-lignified cell walls, Safranin O was used to stain lignified cell walls, nuclei and chloroplast. The results of this are seen in figure 4.5 panel a, in which it is clear that cells in the petal tissue do contain both cytoplasm and nuclei, however in the largest cells it is seen that the nuclei are pressed up against the cell wall, suggesting that a large vacuole is present in the cells. Even at 5 mm, the style is highly lignified, and this lignification continues throughout the development of the flower bud. In figure 4.5b, it is evident that in the petal tissue near the point of anther attachment the cells are quite expanded; the cell contents appear to be pressed against the cell wall, making it hard to determine what is being stained.

Due to the issues caused by these expanded cells, and after numerous attempts, it was decided that the RNA *in situ* localisation in these tissues would not be pursued any further.

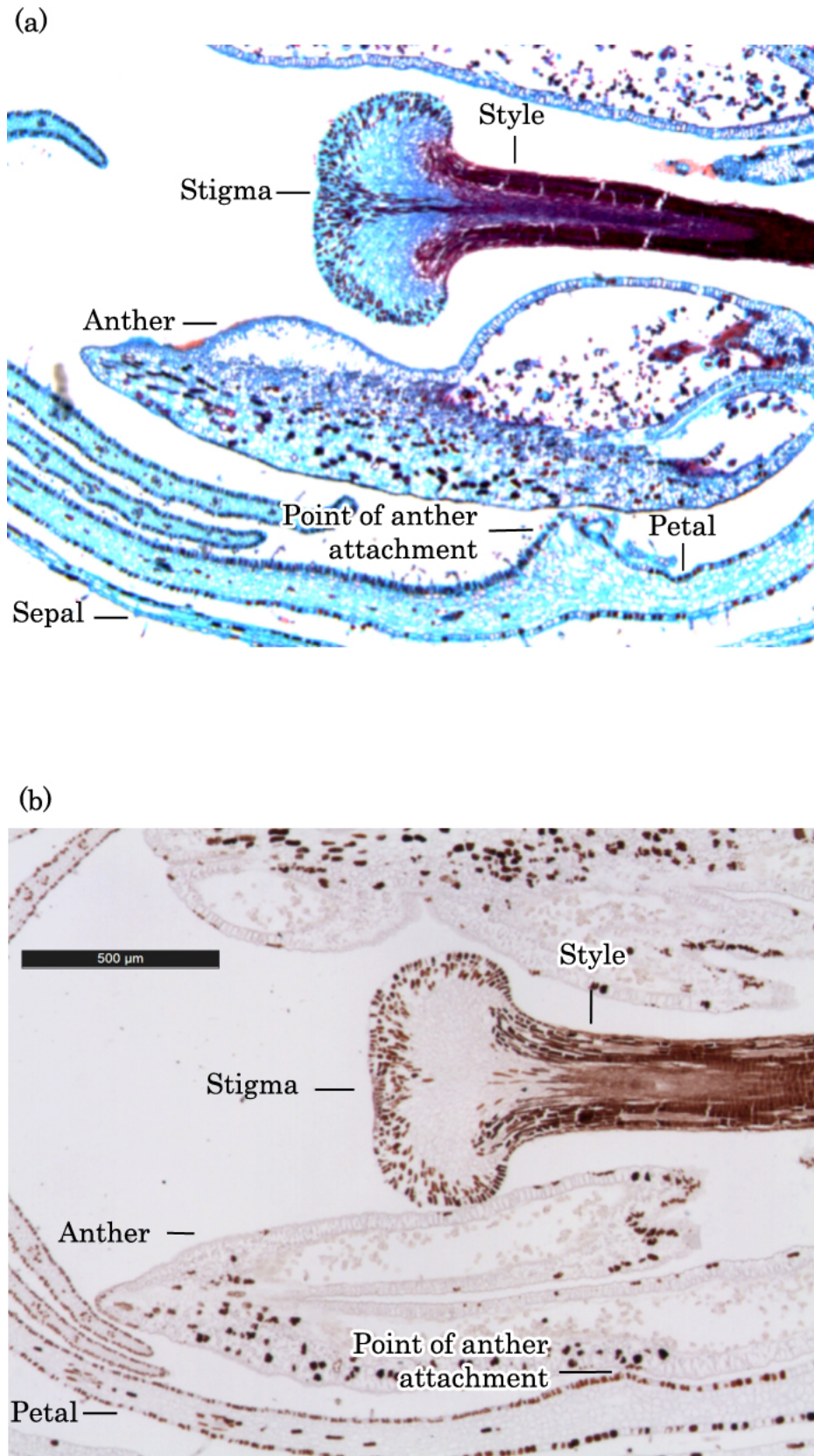


Figure 4.5: Thrum 5 mm flower bud sections. Sections cut at 8 μm . Floral organs are marked. (a) Stained with Alcian blue to indicate non-lignified cell walls, and Safranin O was used to indicate lignified cell walls, nuclei and chloroplasts. (b) Post RNA *in situ* hybridisation.

4.3 Immunolocalisation using peptide antibodies

While RNA *in situ* hybridisation shows localisation of gene expression, it cannot provide any information as to what happens to the protein that is produced. Expression of a gene does not guarantee protein stability, and analysing the localisation of the protein itself could yield information as to its function.

GLO and GLO^T show a high level of sequence similarity, particularly in the MADS box domain; the consequence of this is that antibodies raised to the full-length protein of either GLO or GLO^T would be highly likely to recognise both. As a result, it was decided that peptide antibodies would be used for this investigation. Peptide antibodies are produced by synthesising short stretches of amino acids, which are then injected into animals as would be the case with a full-length protein, and to which antibodies are raised. One downside of peptide antibodies is that, due to the synthetic nature of the peptide produced, they will not necessarily have the same structure as the native protein, and the resultant antibody may not recognise the segment of the native protein it is designed to. The risk of this was, however, deemed to be lower than that of antibodies raised to the full-length proteins showing non-specific localisation.

4.3.1 Design of peptide antibodies for GLO and GLO^T

Peptide antibodies were designed and produced by Dundee Cell Products. Due to the high level of similarity between GLO and GLO^T the identification of even a short region of amino acids unique to one protein was a challenge. When using peptide antibodies, it is preferable to have two peptides designed to the same protein, however this was not possible due to the high sequence similarity seen between GLO and GLO^T, and only one peptide sequence was designed for each protein. As shown in figure 4.6, the peptides were designed to span the C-terminal region of the protein, which shows the lowest level of sequence similarity. In addition to the sequences shown, the GLO peptide has a cysteine residue added to the N-terminal end, and the GLO^T peptide to the C-terminal, for stability reasons; the resultant sequences are GLO: DIEGGEMENGYNC for GLO, and CENKRLGYKFK for GLO^T. Due to 14 amino acid deletion from GLO^T, the GLO peptide sequence was taken from this region, and in doing so guarantees it will not pick up GLO^T.

```

GLO      MGRGKIEIKRIENSNNRQVTYSKRRNGIISKAKEISVLCDAQVSLVIFANSKGMHEYCS
GLOT     MGRGKVEIKRIENSNIHQVTYSNRRNGILKKAKEISVLCDAQVSLIIFSSSSKMHDYCSP
          *****:*****:*****:*****:*****:*****:*****:*****:*****
          :

GLO      KTPLINILDAYQKQSGNRLWDAKHENLSNEIERVKKENDNMQIELRHLKGEDVQSLHHKE
GLOT     NSSLINILDAYQKQSGIRLWDARHENLSNEIERVKKENDNMQIELRYLKGEDIQSLHHKE
          :: *****:*****:*****:*****:*****:*****:*****:*****
          :

GLO      LMSIESALENGLACVRQREMEIYRMARENFADKERVLEDENRSLTYQMHHLVMDIEGEM
GLOT     LMSIEDALENGLTRVRERQMEIYRMAKDNFADKERLLEDENKRLGYKFQQV-----
          *****:*****:*****:*****:*****:*****:*****:*****
          :

GLO      ENGYNYQSQMPFSFRVQPIQPNLQERI
GLOT     -----MDMQMPCSYRVQPLQPNLHDQF
          :  *** *:****:****:***:

```

Figure 4.6: Position of the peptides within the protein sequence for GLO and GLO^T. Comparison between the GLO and GLO^T protein sequences, with the amino acids selected for peptide synthesis highlighted in red. The * indicates conserved residues, the : indicates conservation between groups of strongly similar properties, the . indicates conservation between groups of weakly similar properties, a blank space indicates no conservation.

4.3.2 Design of peptide antibodies for other genes in the *S* locus

At this point in the investigation, a number of other genes had been identified as existing in the hemizygous thrum-only region; *CONSERVED CYSTEINE MOTIF* (*CCM^T*), *CYTOCHROME P450* (*CYP^T*), *PUMILIO-LIKE* (*PUM^T*), and *KISS-ME-DEADLY KELCH REPEAT F-BOX* (*KFB^T*). One of these recently identified thrum-specific genes, *KFB^T*, has no paralogues. As a result, designing a peptide for this protein was an easier task; the chosen peptide sequence was EVIPGLPEDLGLE, and this did not need any modifications. The position of the peptide within the protein sequence is shown in figure 4.7.

Another gene found within the genome has been designated *PUMILIO-LIKE* (*PUM^T*) due to the high level of sequence similarity it displays with the *PUMILIO* genes of other species. The exact function of this gene in *P. vulgaris* is not yet known, however its existence within the hemizygous region would suggest that it, along with *GLO^T*, may be involved in one of the functions of heterostyly. The region of sequence selected for antibody production was CKGSRGRKRRSHKK, and an additional cysteine residue was added to the C-terminus for stability.

There are two further genes present in the *S* locus for which peptide sequences were intended to be selected: *CCM^T* and *CYP^T*. *CCM^T* shows high similarity to another protein, *CCM^T-LIKE*. These two proteins show a very high level of sequence similarity; the percentage similarity is 83%, and the longest contiguous stretch of differences between the amino acid sequences is two. As a result, it has been decided that at this time no peptides will be synthesised for this gene, and no antibodies will be raised due to the high risk of the antibody recognising the paralogue.

CYP^T has been found to be part of a small gene family with a number of closely-related paralogues. There are very few stretches within the sequence of *CYP^T* that differ from all of the paralogues. As a result, it was decided that, as the antibody designed to *CYP^T* would also recognise at least one of the paralogues, it would not, at this stage in the investigation, be suitable for use in this experiment.

MEVIPGLPEDLGLECMIRSHYTTFRVVSQTCHLWRKLLQTTFHSYRKNKGYSHKMICFV	60
QSIPPNALADETGKSANSCGYGITVFDLRSRTWGRLSQVPKYQSGPLPLFCRVASSGNKLI	120
VMGGWDPFSYHPVKDVFVYDFVNQLWRQKDMPSKRSFFAMGAIDGHVYVAGGHDENKGA	180
LKSAWFMIDGMSGLR	196

Figure 4.7: Position of the peptide within the protein sequence of KFB^T. The amino acids selected for peptide synthesis highlighted in red.

MFSSFESTSSTISSDVPAAIAQMDSIIHHRMTNHIVDFCKESTDWIQHEMDICDDL SRIN	60
IGVPHKHYSKALNFNGFASRYCEDYSKADYLSPHVLGSI ESYCKNKL TNFECCTKNKNSL	120
NCEDREQQYYDLTVNKPCLDQDLSYLEHCILKSVGDVNPLNLSKLFRFVYPNQFLSVDTN	180
HICTGSLSYSSQRLVEAINTVCHGSVLYPNSASSERKSTVNCSVFLNSCADTYSERNGFM	240
VYINQATILELFLNNISSNRRSSVENPICIDSMIYVNRSLNVLCVSNHLMFYSNRVSRE	300
DVVLYNVSMMPN KDSSLSQLCNCMLFRDEIRSHKNGRAILGRVNF CCKHSF ILEGQSL	360
RLVAEKQLEHGKGSRGKR RSHKNMSLKCNGRTMTRLIQNLMCKKSMVSQIIRGTLLE	420
FRGYVYLI AKNQIGCLFLQMI LEEGNHHD IQVIFNETINHMAELIMDPFANH LIQKLLSV	480
CGDEQITQIVLKVTAKHGRFITICFNTHGTRVLQKLIKSLNTRQQQKL VVSALQRRFLEL	540
VKDENG YHIIK SCLQAAPKCRVGYI LKSCIAQSVGKFREKMVTEIASEGFDLAQDAFGN	600
YVIQHIIELNIPSAAAILSSQFHGNYVYLSTQKFSSYVVQTF LKSYKQSLPIIIQELL SV	660
PHFDHLLKDPFANYVIQTAVDISKVLIRPSFYC	693

Figure 4.8: Position of the peptide within the protein sequence of PUM^T. The amino acids selected for peptide synthesis highlighted in red.

4.3.3 *In silico* testing of peptide specificity

In addition to the protein sequences and alignments seen in figures 4.6, 4.7 and 4.8, the chosen peptide sequences were compared to the translated protein sequences of the *P. vulgaris* gene models, by way of a PBLAST alignment, with parameters optimised to detect similarity despite the short input sequence. This was conducted to determine whether these sequences were present in any other *P. vulgaris* proteins, as a high level of similarity between the peptide sequence and the translated protein sequence of another *P. vulgaris* gene would result in off-target binding of the antibody when used in immunolocalisation experiments. This analysis was performed by Jonanthan Cocker. Other than the intended target, all peptides do not show high levels of similarity with other protein sequences, and as a result, it is likely that they will only recognise the intended target. The sequences of all of the synthesised peptides, including C- and N-terminal cystine additions, are shown in table 4.1.

Due to time constraints these peptides were not tested as part of this investigation; their future uses and the implications of the data they may generate are discussed in Chapter 6.

Table 4.1: Peptide sequences used for antibody production. Sequences are shown C- to N-terminal. Additional cystine residues added for stability are highlighted in red.

Protein name	Peptide sequence
GLO	DIEGGEMENGYNC
GLO ^T	CENKRLGYKFQ
KFB ^T	EVIPGLPEDLGLE
PUM ^T	CKGSRGRKRRSHKK

4.4 Summary of findings

At the point of floral organ determination, the expression profile of *GLO* and *GLO^T* appears to differ. *GLO* is involved in specifying the identity of floral whorls two and three, as it does in its *A. majus* orthologue (Tröbner *et al.* 1992), and as such is tightly confined to these areas as the floral whorls begin to differentiate. Expression of *GLO^T* does not show such well-defined localisation, and has been shown to be expressed throughout the floral meristem. In Chapter 3, quantitative real-time PCR (qPCR) data revealed expression of *GLO* and *GLO^T* continues into the adult flower bud, and semi-quantitative PCR has localised them to whorls two and three, however RNA *in situ* localisation has failed to corroborate these data. In this tissue and under these conditions it would appear that RNA *in situ* hybridisation is not a suitable technique for localising expression.

While it has not been possible to test the specificity of the peptide antibodies that have been raised to the genes at the *S* locus, their production and potential future use in immunolocalisation experiments will provide further information as to the localisation of these two proteins, and how they may or may not have diverged.

CHAPTER 5

Expressing *Primula vulgaris* MADS box genes in *Arabidopsis thaliana*

5.1 Introduction

The development of floral architecture in angiosperms has been shown to be the under the control of a collection of genes acting in unison; the method by which they act is referred to as the ABCE model (Bowman *et al.* 1991; Causier *et al.* 2010). In this model, the expression of genes in tightly controlled regions of the floral meristem brings about the specification of the four whorls of the flower: sepals, petals, stamen, and carpels (Bowman *et al.* 1991).

The majority of the work that led to the proposal of this model has been done in *Antirrhinum majus* and *A. thaliana* (Bowman *et al.* 1991; Coen and Meyerowitz 1991). Homeotic mutants, in which the gene expression of floral organ identity genes is altered, resulting in unusual floral phenotypes, paved the way for the creation of this model (Meyerowitz *et al.* 1989). There are three main categories of gene that specify organ identity: these are referred to as A, B and C function genes, and the combinations in which these genes are expressed determine which floral organ is generated. The expression of A function genes alone results in sepals, a combination of A function and B function results in petals, a combination of B function and C function in stamens, and C function alone in the formation of carpels (Bowman *et al.* 1991). The more recently discovered E function *SEPALLATA* genes, act in protein

complexes with the A, B and C function genes to specify organ identity, leading to the suggestion that the model should be renamed the ABCE model (Causier *et al.* 2010).

In *P. vulgaris*, there are two known B function MADS box genes, *GLOBOSA* (*GLO*) and *DEFICIENS* (*DEF*), which, when dimerised specify the organ identity of the second floral whorl (Li *et al.* 2008). *GLO* is an orthologue of the *A. majus* B function gene *GLOBOSA* (*GLO*) and the *A. thaliana* gene *PISTILLATA* (*PI*) (Tröbner *et al.* 1992; Goto and Meyerowitz, 1994). The thrum-specific duplication of *GLO*, *GLO^T*, shows a different pattern of expression through the growth of the flower bud; in Chapter 3 it is shown that *GLO* expression increases through development, whereas *GLO^T* remains constant, and in Chapter 4 *GLO* shows a distinct pattern of localisation in the second and third whorls of developing floral meristems, whereas *GLO^T* expression is less well defined.

The proposed function of *GLO^T* is as the gene responsible for elevating the anthers in the thrum form of *P. vulgaris*. Plants displaying the short homostyle phenotype, in which both the anthers and the style are positioned half way down the corolla tube, have been shown to contain a transposon insertion in the coding sequence of *GLO^T* (Li *et al.* 2016). As of yet, little is known about how it brings about the function of raising the anthers in thrum plants, or how its function has diverged from that of *GLO* since the duplication event. It has been shown that in the *Hose-in-Hose* mutant, in which an insertion into the promoter of *GLO* results in ectopic expression and subsequently the transition of sepals to petals, expression of *GLO^T* remains unaltered, which may suggest its role has diversified (Li *et al.* 2010). Analysing the function of this gene would best be achieved by over-expression and the generation of knockout or knock-down plants within the species of interest. In the case of *P. vulgaris* regeneration and transformation studies are in their infancy (Hayta *et al.* 2016), and as such these experiments are not yet possible.

5.1.1 Experimental aims

The ease of transformation of *A. thaliana* via the floral dip method (Clough and Bent 1998) makes it an excellent candidate for assessing the functional divergence of *GLO* and *GLO^T*. While *A. thaliana* does not display heterostyly or self-incompatibility this does not prevent the assessment of whether *GLO* and *GLO^T* are still performing similar functions.

The design and construction of vectors in which the promoters of the *P. vulgaris* genes were fused to a fluorescent reporter, and their subsequent introduction into *A. thaliana* via the floral dip method, would add to what is known about the localisation of *GLO^T* in comparison to *GLO*. In tandem, *in silico* analysis of the promoters would show divergence between regulatory elements controlling the expression of the two genes. As a result, promoter-reporter constructs will be generated for these genes that can be used in *A. thaliana*.

More information is needed as to how diverged the *P. vulgaris* genes are, and how *GLO^T* functions in comparison to *GLO* will yield valuable information as to the function of *GLO^T*. Yeast 2-Hybrid (Y2H) interaction data, in which the *GLO*, *GLO^T*, *PI*, and *APETALA3* (*AP3*) proteins will be expressed in Activation Domain (AD) and DNA Binding Domain (BD) fusion constructs in yeast, will be generated to test protein-protein interactions. All combinations of AD-fusion and BD-fusion will be tested, and these data will provide information that can inform future expression experiments.

In addition, constructs in which the *GLO* and *GLO^T* coding regions were driven by the native *PI* promoter were introduced to the *A. thaliana pistillata-1* (*pi-1*) mutant. The intention of these experiments was to further expand knowledge as to how diverged *GLO^T* is from *GLO*.

Dr. Sadiye Hayta has achieved the overexpression of *GLO^T* in *A. thaliana*, and the overexpression of *GLO^T* results in a different phenotype to the overexpression of *PI*, suggesting that the genes perform different functions. Overexpressing *PI* in *A. thaliana* under the control of the 35S promoter of the Cauliflower Mosaic Virus (CaMV) has been shown to result in the partial conversion of the first whorl from sepals to petals; the tissue is a mosaic of the two, with cells at the base of the organ resembling those of petals, and those in the upper regions resembling sepals (Krizek and Meyerowitz 1996). Plants in which *GLO^T* is overexpressed do not show the distinct first whorl-only phenotype of *PI* overexpression, and instead show a range of floral phenotypes. A number of the *GLO^T* overexpression plants appear similar to, but do not directly mirror, the *cauliflower* mutant (data not shown), in which there is a complete breakdown in the floral architecture (Bowman *et al.* 1993).

The creation of a *GLO* overexpression vector for use in *A. thaliana* had not yet been accomplished, and the production of this vector and its introduction into *A. thaliana* would provide a further point at which the expression patterns of these two genes can be compared.

5.2 Promoter-reporter constructs design for *GLO*, *GLO^T* and *PI*

At the point of this investigation, it was not known how long the full promoter sequences of *GLO* and *GLO^T* were, and how far upstream of the transcription start site would have to be included for the promoter to still function. It has been shown previously that in *A. thaliana* 1.5 kb of the *PI* promoter is required for full functionality (Honma and Goto 2000), and while this is not a direct indication of how long the promoter is likely to be in *GLO* and *GLO^T* it was decided that 2 kb upstream of the start codon would be used. Including too little promoter could result in a non-functional construct, or one that results in different spatial or localisation to that seen in wild type *P. vulgaris*, as is seen in promoter deletion experiments conducted on *PI* (Honma and Goto 2000).

In addition to the *P. vulgaris* promoters, the *PI* promoter will also be used. It is possible that, due to sequence differences, the *GLO* and *GLO^T* promoters may not confer an authentic expression profile in *A. thaliana*, and using these alone would result it being unknown if the experiment failed due to the construct or the promoter sequence being too dissimilar to that of *PI* in *A. thaliana*. As the localisation of *PI* expression in *A. thaliana* has been well documented (Goto and Meyerowitz, 1994), a construct containing the *A. thaliana PI* promoter will be used as a positive control.

5.2.1 Construct design for promoter-reporter experiments

The *Primula* genome browser was used to obtain the contigs in which *GLO* and *GLO^T* are found, and using these, the required sequences were selected for *GLO* and *GLO^T*. 2 kb upstream of the *PI* start codon was also used. As the *pi-1* mutant is in the Landsberg *erecta* (Ler) *A. thaliana* ecotype, this ecotype was used in these experiments.

Type IIS restriction enzyme cloning was used to generate these constructs, and the enzyme used in this investigation was BsaI. This class of restriction enzymes cuts beyond the recognition site. This means that two PCR products can be designed so that, when cut with BsaI, the overhangs can be ligated together seamlessly; this process is described in section 2.8.9.1 and figure 2.1.

Due to the method of construct assembly, all *P. vulgaris* sequences were searched for BsaI sites using RestrictionMapper (<http://www.restrictionmapper.org>). The *GLO* and *GLO^T* promoters contained no BsaI sites, however one site was present in

the promoter of *PI*. As a result, two versions of the *PI* promoter were created, in which the BsaI site had been removed by splitting the promoter into two parts, and using primer sequences to change one base of the recognition site, as described in section 2.8.91 and figure 2.2. In the first version, the first base of the BsaI recognition site (GGTCTC) was altered, and in the second version the last base was altered. Two versions were created in the hope that, if a transcription factor binding site was present within this region, only one of the version of the promoter would be rendered non-functional.

The fluorescent reporter chosen for these experiments was GFP, as it has been used extensively in the analysis of gene expression in *A. thaliana* (Hanson and Köhler 2001). Vectors containing the coding region of GFP (pICH41531) and the NOS terminator (pICH41421) were obtained from The Sainsbury Lab (Engler *et al.* 2014). The required sequences are released from the vectors by digestion with BsaI. The binary vector (pICSL86900) was also obtained from The Sainsbury Lab, and allows kanamycin selection in plants. When digested with BsaI the binary vector leaves known overhangs to which the first and last constituent part of the constructs match. Plasmid maps of pICH41531, pICH41421 and pICSL86900 are available in the appendix, and figure 5.1 shows how the constituent parts will be arranged within the binary vector.

An advantage of this method is that the final construct assembly can be conducted as one digestion and ligation reaction; when these constructs are digested with BsaI, they will result in overhangs that will only ligate in the designated order. Additionally, it allows for fragments to be joined seamlessly, reducing the possibility of important binding sites within the promoter being disrupted.

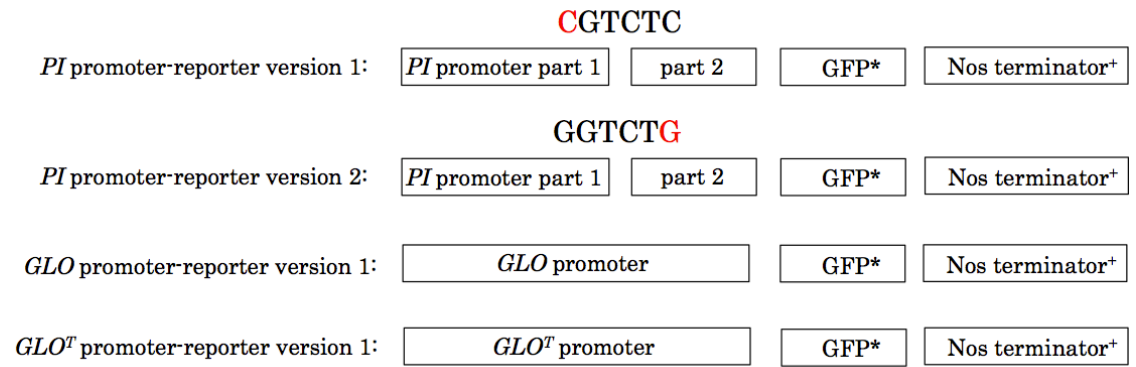


Figure 5.1: The four promoter-reporter constructs. Constituent parts of the four promoter-reporter constructs in the order they will be arranged when transferred to the binary vector. The two versions of the *PI* promoter are indicated; sequences above the boxes indicates the BsaI recognition site, and the base that has been changed to remove the BsaI site is highlighted in red. *GFP = pICH41531 from The Sainsbury Lab, ⁺pICH41421 NOS terminator from The Sainsbury Lab.

5.2.2 Promoter-reporter construct production

The necessary promoter sequences were amplified from *P. vulgaris* genomic DNA by PCR, which was conducted using primers with the BsaI restriction site added to their 5' ends; these sequences are available in the appendix. These PCR products were cloned into pGEM-T Easy and verified via single read sequencing before use. The vector or vectors containing the promoter sequence, and the vectors containing GFP coding region and NOS terminator and the binary vector were then digested and ligated (section 2.8.9.2).

The binary vector was introduced into DH5 α *E. coli* cells, and selected for using kanamycin. PCR was used to check that the fragments had ligated in the correct order, with primers designed to amplify the joined regions. Whilst the overhangs are designed to be specific and should ensure fragments ligate in only one order, it is possible that they may mis-ligate. An example image of this process is shown in figure 5.2 (a). Following successful PCR amplification, the joins in the plasmid were subsequently verified by single read sequencing; an example is shown in figure 5.2 (b). The method of verifying the order of the constituent parts of each construct by PCR was repeated when the construct had been transformed into *A. tumefaciens*; the results are shown in 5.2 (c).

Once assembled into the final binary vector, these constructs were introduced into *A. thaliana* Ler plants. The Ler ecotype was used, as it is the ecotype in which the *pi-1* mutant is found, and hence, the gene has received the most previous study. The *A. tumefaciens* strain for *A. thaliana* transformations of the complementation vectors was GV3101 (section 2.8.9.5). Full details of the floral dip procedure are available in section 2.9.

5.2.3 Analysis of transformants for promoter-reporter constructs

The transformant offspring produced in this experiment have not yet been fully investigated; the phenotype expected will only be found in the floral tissues, and there was not enough time for these analyses to be carried out within the time constraints of this project. The potential for future work regarding these transformants is discussed in Chapter 6.

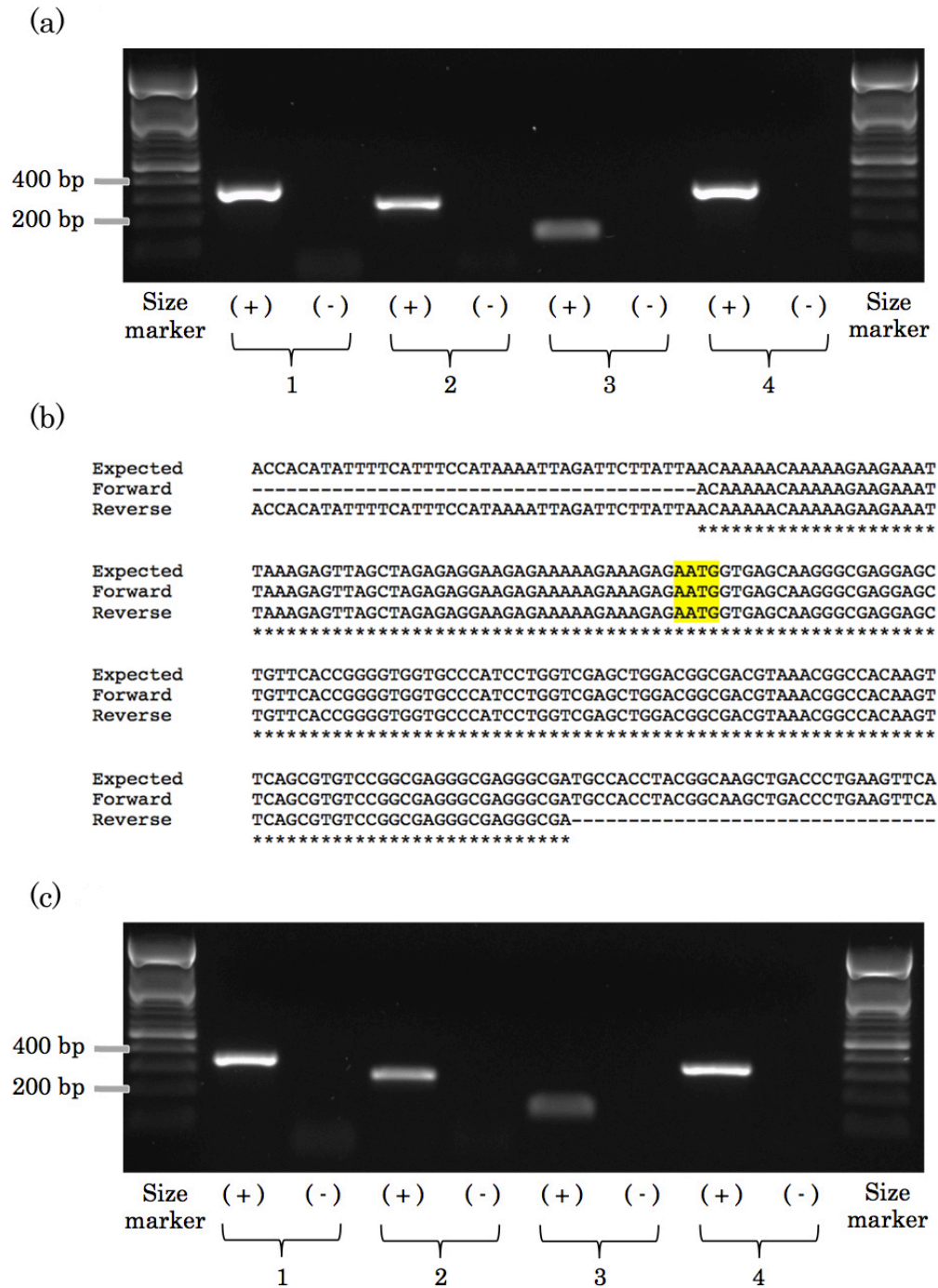


Figure 5.2: Example of the analyses performed on the promoter-reporter constructs.

(a) Standard GoTaq PCR conducted on *GLO* promoter-report plasmid DNA using primers designed to verify that constituent parts of the construct have ligated in the correct order. Joins being tested are: 1 = binary vector and *GLO* promoter, 2 = *GLO* promoter and GFP coding region, 3 = GFP coding region and NOS terminator, 4 = NOS terminator and binary vector. (+) indicates plasmid DNA, (-) indicates no DNA control. (b) DNA sequence alignment of two joined fragments. Expected sequence was compared to sequence data in the forward and reverse orientations. The example shown is combination 2 from panel (a); the joined overhang is highlighted in yellow. (c) Standard GoTaq PCR conducted on *A. tumefaciens* containing the *GLO* promoter-report plasmid, using the primer combinations from panel (a).

5.3 *In silico* promoter analysis of *GLO* and *GLO^T*

An *in silico* comparison of the promoter sequences of the two genes was conducted in the hope of establishing to what degree the promoter sequences have differentiated since the duplication event that created *GLO^T*. The promoter sequences of *GLO* and *GLO^T* are aligned in figure 5.3. While there are some regions in which the sequences are similar, there are clear differences between the two. In *A. thaliana*, it is known that, in the later stages of development, PI and AP3 form a dimer and maintain the expression of *PI* via a proximal promoter (Honma and Goto 2000). While the position of the transcription factor binding sites may not be the same within the *P. vulgaris* promoter sequences, the auto-regulation of *GLO*, by the *GLO-DEF* dimer, is very likely. Whether *GLO^T* contains these same transcription factor binding sites will further inform us as to how different these genes are. As a result, this search of the promoters will focus specifically on MADS box transcription factor binding sites.

GLOT	-TTTGAGCTTTTTCTCACACAATGGTGATTTTTAGGGTTATAATGGAGAGAGGTTGAAGG	59
GLO	CTTACTCATTGATGTAAC-----GCATCTAGCTATACGTATAATTGTGAAATTTTGAAGT	55
	* * * * *	
GLOT	ATGG-----GTGAATTGTTTTATTTGAAGAGGTAATTGGGGGGTATATTAAGGA	109
GLO	AAATCTCAATCAAGGCACCCATTATAACCAAAATTTGTGAATTCACAAAAAAGAGTC	115
	* * * * *	
GLOT	-----AATAAGGTTTGTACCGTTATCTAAAGGTTTATGTCTTTACCCA-----	154
GLO	ACTTCAGTATGCGGCTTGGCATTACAATTTCCAAAGGTTTCAGTTTTTGAACCATGCAAC	175
	* * * * *	
GLOT	-----TTGGTTTAAAGCAATTTCCCAATTTTCCAGAATAAGTCGAAAAGCTGTTTT--T	206
GLO	ATCTAATTGGACTGCATACATAACACTCTCACTGAATCAAAGCAAACCTAGCAATAT	235
	* * * * *	
GLOT	TTGAATTCTGCAGAAAACGATCGGTCGTTCTCTGAAAAACGATCGGTCGTTTTCTGAAAA	266
GLO	TATAATTCTAGACTCAAAGTCC-----TACAAATAGATATCAAATCAAGGTTTCTACTTC	290
	* * * * *	
GLOT	ACGATCGGTCGTTTTCTGAAAAACGATCGGTCGTTTTCTCAAAAACGATCGGTCGTTTTC	326
GLO	CTCTTCACTCTCTTGA AAAATGACAGTAAATATGGTTTTTAAAAATGGGAAAATAATATT	350
	* * * * *	
GLOT	TCAAAAACGATCGGGTGTTTTTATCATGAACCTACTGGAATATTTTGGAAACAAAACGAT	386
GLO	GCAAAAACGTAAAA-----GATAGTATTACACTTCTTTGTCCAAAACCTCCAA	398
	* * * * *	
GLOT	CGGTCGT-TTTGGTCATTTTCGATCGGTCGTTTTTTCAGACGACCTTACTGGAACATTTTGG	445
GLO	ACCTTTTTTTAGATGTTTCTCCTTTAT-----TATGTGGGAAGTGACATATATTTG	448
	* * * * *	
GLOT	ACAAAAACGATCGGTCGTTTTGTCCAAAACGATCGGTCGTTTGGTCCAAAACG----AT	501
GLO	GATATGATGGTTAAGG---TTTGATGAACACACTTGTT-TTTTTTGTCTATCCTTTAAAT	504
	* * * * *	
GLOT	CGGTCGTTTGTCTGTAAATTTCTCGGTTTTCCGCTATACTTCGAATGCTAGTAACTTCT	561
GLO	CAAAGACATGA-CCCTCCAAACTCCAATGTGCCTAGCACTAGCTTCTACTAGCTTAA	563
	* * * * *	
GLOT	TAGTTTTAGCTCCGATTCTAGTTCGCTTGGACCGTTGGCTTCATATTTGATTACTCTAT	621
GLO	TAACGTAT----CAAAGATTGATCCTCGGTATAGAGGGGTACAACTTTAAACCTTCTCA	619
	* * * * *	
GLOT	----CTTCTGATGTTCTGAAATTTACATTTTCTCTTTTGGAAAAGACAAATCA-----	672
GLO	AATTTTATTTAGTGTTTTGGCTGTTACTATGTAACATCAAATAACCCAAAAACGTGTT	679
	* * * * *	
GLOT	-----TTATATCATCGAATAGAGAAATCGTAAAATCATTAA-----TACAT	713
GLO	TTTCTTTTACCACCTTTTCGTAATAATGTAAGTACATTAAACAACTATTATTTATAACT	739
	* * * * *	
GLOT	AATCAAACGTATAATCAAAATCAATTATTGAACCTATGGTTCAACACTCCCAATTTGATA	773
GLO	TCTATGAGGTATAACAAAATTATATTATATAACGTTTTGTACAAATTAGGATGCTGTGCA	799
	* * * * *	
GLOT	TAAAAATCAAACTAATTGGTCTGAATCTCTAAATCGAAGTTTTTCTCTAAAATCACCAT	833
GLO	---AATGTTAA----TATATATGTATCATGGTCACCAATATTGACAGTATATATGCCCGC	852
	* * * * *	
GLOT	AGTAACAAAATCCGAGTAAATC-----ATCACATCATCTATGGAGATGCGGAT	881
GLO	TTCCGTCTAATCATGTTTAATCTTGTACGATCATGTCTAATTATGTGTGGTTCTGTGCT	912
	* * * * *	
GLOT	GGATAGGGGGAGATAGATAGATATAGATAGGTATATATAGGGAGAGTATATAAAAGAGA	941
GLO	GCCTGGTTGCGTG-----TCTGGTTTTTTTTTTTTTTGGGTATTTTTTTTCACTTT	964
	* * * * *	
GLOT	AGGGATATTTTTTGGAGAGAGAGAGAGAGTGGGAGTTTCAGATTTGTATTTTTGAGGATC	1001
GLO	TTTAGTTTTTTTTTACCA-TTTTTTTGTGTTGG-----TATTTTTTTTTTCACTTTT	1014
	* * * * *	
GLOT	GTAGTTTTTGA AAAACAATTATTTGTAA-----TAACTTTTGGTAATTAGAGTTGAAA	1056
GLO	TTAGTTTTTTTTTACCATTTTTTGTGTTTTTTTGGACTTTTCCAGAATTTTTTTAGCGA	1074
	* * * * *	
GLOT	AGTTGTATTTATTG-----GTAATATCTCTTTTTGTAAAT---GATTAGCAAGTGTT	1106

GLO	TTT TAGATACTTTTCAAAGTTGTTATTTCTTTTAGTTTTTTCATTTTTGGCGGTTTTG	1134
	* * * * *	
GLOT	AC-----TAAACTTCAATTAATAATTTGAAGTG--ATAGCTTCTTTTAAA--TAACTA	1156
GLO	ATTTTTTTTGATTTTTTTTATAATGTTTCCATGTTATATATTTTTTGGTGTTTTAATA	1194
	* * * * *	
GLOT	TTTTTAAATGGTACCGACTAAAACTCACTCTAATTTCGGGAACATGGACAACATATTATG	1216
GLO	CATTACAATTAATAAAAAGACGAAAACCGAAATATTTAGAAAAATGGACTAAAACCAA	1254
	* * * * *	
GLOT	AGACATTTTATTGTCAAACAATTTACTCAATTAGGTTATT-ATT-TTTGGGAACAACATA	1274
GLO	ATAAAGTTCATAAACGAGACTCGTCTCCAATTGCCCTAATTAATTGTATAATGGCAACTT	1314
	* * * * *	
GLOT	T-TTTGAAATTGACTTTTTGGTAATTAGAGTTGAAAAGTTTtagttattggtaatatctt	1333
GLO	GATTTTCTTTTGATTTTCCATTGATTATTTTGTGCAAGTTC-----GAGGTGTGTTTTT	1368
	* * * * *	
GLOT	CTTTTTTAAATGATGAGAAAGTGTTACTAAACTCCAATTAATACTTTCAAGTGGTAAC	1393
GLO	AGTGGCACAAAAGCAATTTTAGTGTCCCAACTCACCATTAAATCTGGC-----	1418
	* * * * *	
GLOT	TTGTATTAAATAAATATTTTCAAATGGTATCGGCTAAAACTCACTCTAAT-----	1444
GLO	-----AGTTCACTTAATATGTTTCAGCTACAAACGTGCAATTGTTATAACAG	1467
	* * * * *	
GLOT	-----TTATGGAACATGGACAA--CTATTATGAGACATATTATTGTC	1484
GLO	ATAATGTTACACAGTAGATGTTGGAACAAGAACTTAGTAATCCCAGAATGAAATTTATC	1527
	* * * * *	
GLOT	AAACAGTTTACTCAATTGGGTATTATTATTTTATATAATCATGGGATAAGAACTAAA	1544
GLO	AAGAAATCAATATCATAAACTTAAATAATACTGTACTTGCTATATATACTGATCAAAT	1587
	* * * * *	
GLOT	TACTCCTTGCTAATACGAAATTGGTTATTGATATAAGTAGAAAAATAAACGCTACAGAT	1604
GLO	GTCTCATAACTAGTAC-----TGCTATTGGTAAGCCACACACAAGCCATTA	1635
	* * * * *	
GLOT	ACAGATGCACAGGTAGA--AGTGGGTATCATGGTTGAAATAGTGAAGCTTGGGTTAA	1662
GLO	ACCAATAGCAAAAATGATGAAGGAACCCAGTGGTTAAAAGATTGAATGGCTAAACAAG	1695
	* * * * *	
GLOT	AGACATATAAAAAACCTCTAAAAACATATAAACTTATTATTGCAATAAATAAGATATA	1722
GLO	GTACTCAAAGTATAAAATAAAA-AATACTACAAGTACCATTAAAGTAGAAATAAATTCA	1754
	* * * * *	
GLOT	AAGTTAAGATGACACAACAGTGTTCAGTAGGATTGGTTTGTCA-AAACCTTAAACA	1781
GLO	AAGTT---GAAATAGTAAAGTGTTCAGTAGGTTGGCATTGACAAGTGCCTTAAACA	1811
	* * * * *	
GLOT	AACATTGCATGCTGTGACTCTACTAAAGTCTCTTGCTCCAACCTCTAATGGAACTAC	1841
GLO	TCCCCTCTAAGCTGTAACCTCTTAAAGTCTCTTGCTTCAACCTTTAATGGAACTCT	1871
	* * * * *	
GLOT	TCCCTTTCCCTCTCTCTGTATCTTCAGCAAATCTTCTCATTACTCGATGCCATTATAT	1901
GLO	TACT-TTCTTTCTCTATCTATCTTGACCACATATTTTCATTTCCATAAAATTAGATT-	1929
	* * * * *	
GLOT	CAACAAGAACAAAAATCATATCTACGTTTATAAAATTATTAAAGGGAAATAAATAGAGAA	1961
GLO	-CTTATTAACAAAAACAAAAG-----AAGAAATTAAGAGTTAG	1968
	* * * * *	
GLOT	CAAGAAAGCTAGAGAGATAAGAAGGAAGATG	1992
GLO	CTAGAGAGGAAGAGAAAAAG--AAAGAGATG	1997
	* * * * *	

Figure 5.3: Alignment of the *GLO* and *GLO^T* promoter sequences.

The MatInspector module (Cartharius *et al.* 2005) of the Genomatrix Software Suite (v3.7) was used to identify potential MADS box transcription factor binding sites from data collected from numerous plant families, available via the Matrix Library 9.4. For both *GLO* and *GLO^T*, 2 kb of sequence upstream of the start codon was compared to the database of known transcription factor binding sites; the results of these searches can be found in figure 5.4 and table 5.1 for *GLO*, and figure 5.5 and table 5.2 for *GLO^T*. Each site that is found is given a ‘Matrix similarity’ value; transcription factor binding site are considered a ‘good’ match to the matrix if the value is >0.80 (Cartharius *et al.* 2005).

As shown in tables 5.1 and 5.2, both genes contain a number of putative MADS box transcription factor binding sites in their promoters. *GLO* contains more in total, and is predicted to have two binding sites for the AP3/PI dimer, one of which is 800 bp away from the start codon; it is possible that this transcription factor binding site may perform the same function as the proximal promoter region responsible for sustained *PI* expression (Honma and Goto 2000). There are none of these AP3/PI dimer transcription factor binding sites present in *GLO* although it is predicted to have a number of transcription factor binding sites at which other floral organ determination genes could bind.

The downside of this method of analysis is that it is purely theoretical; any findings must first be verified before conclusions can be drawn. However, the differences in the promoter sequences would suggest some changes in regulation have occurred, and this is supported by the evidence from Chapter 3 and Chapter 4, in which *GLO* and *GLO^T* show differences in their spatial and temporal localisation.

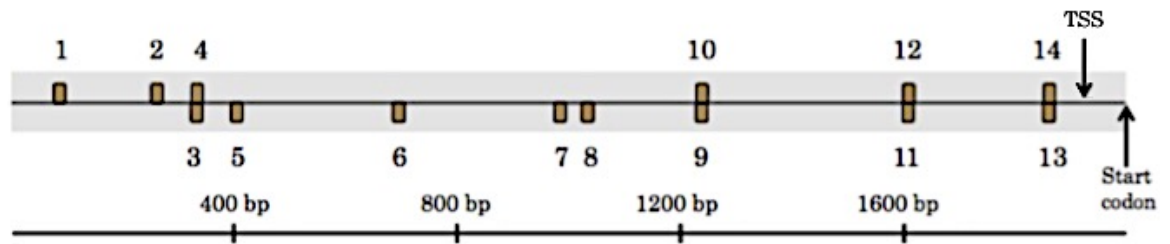


Figure 5.4: MADS box transcription factor binding sites within *GLO* promoter. Markers above the line indicate the binding site was found on the (+) strand, markers below the line indicate the binding site was found on the (-) strand. Transcription start site and ATG are indicated. Numbers correspond to the matches in table 5.1.

Table 5.1: Description of the MADS box protein binding sites within the *GLO* promoter. As shown in figure 5.2, with start and end positions listed for each. Lower numbers indicate the binding site is further from the TSS and ATG.

No. on figure	Description	Start position	End position	Matrix similarity value
1	MADS box protein AGAMOUS	78	98	0.829
2	MADS box protein SQUAMOSA	252	272	0.956
3	Binding sites for AP1, AP3-PI and AG dimers	323	343	0.844
4	MADS box protein SQUAMOSA	324	344	0.916
5	MADS box protein SQUAMOSA	395	415	0.909
6	MADS box protein AGAMOUS	685	705	0.817
7	MADS box protein SQUAMOSA	974	994	0.902
8	MADS box protein SQUAMOSA	1023	1043	0.902
9	Binding sites for AP1, AP3-PI and AG dimers	1227	1247	0.792
10	MADS box protein SQUAMOSA	1228	1248	0.902
11	AGL2, Arabidopsis MADS-domain protein AGAMOUS-like 2	1597	1617	0.83
12	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15	1598	1618	0.791
13	AGL3, MADS Box protein	1849	1869	0.86
14	AGL2, Arabidopsis MADS-domain protein AGAMOUS-like 2	1850	1870	0.857

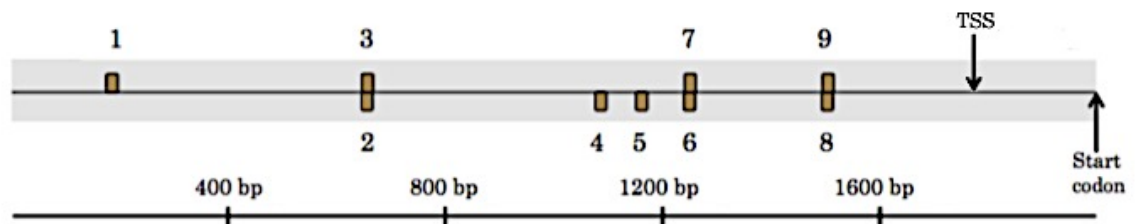


Figure 5.5: MADS box transcription factor binding sites within *GLO^T* promoter. Markers above the line indicate the binding site was found on the (+) strand, markers below the line indicate the binding site was found on the (-) strand. Transcription start site and ATG are indicated. Numbers correspond to the matches in table 5.2.

Table 5.2: Description of the MADS box protein binding sites within *GLO^T* promoter. As shown in figure 5.3, with start and end positions listed for each. Lower numbers indicate the binding site is further from the TSS and ATG.

No. on figure	Description	Start position	End position	Matrix similarity value
1	AGL3, MADS box protein	176	196	0.831
2	Flowering locus C	644	664	0.971
3	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15	645	665	0.851
4	Tomato MADS box transcription factor MADS-RIN	1073	1093	0.777
5	MADS box protein SQUAMOSA	1148	1168	0.939
6	MADS box protein AGAMOUS	1236	1256	0.834
7	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15	1237	1257	0.805
8	AGL2, Arabidopsis MADS-domain protein AGAMOUS-like 2	1489	1509	0.831
9	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15	1490	1510	0.801

5.4 The interactions of *P. vulgaris* and *A. thaliana* MADS box transcription factors

The level to which *GLO^T* has diverged from *GLO* in terms of its function remains, as of yet, unknown. The difficulties involved in transforming a species such as *P. vulgaris*, combined with its bi-annual flowering pattern, mean that functional analysis by way of overexpression, knock-out and knock-down experiments involving the manipulation floral genes are not yet easily achievable. As a result, analysing the function of these genes in easily transformable species may provide useful information as to how their functionality differs.

5.4.1 The interaction of *P. vulgaris* and *A. thaliana* genes in yeast

The Yeast 2-Hybrid (Y2H) system allows protein-protein interactions to be investigated. The Y2H system is based on the premise that a reporter gene requires a functional transcription factor to bind to it and activate its expression, and that a transcription factor must consist of an activation domain (AD) and a DNA binding domain (BD) for this to be possible. The sequences of interest are cloned into two different vectors, one which contains the AD, and one which contains the BD. If, in the yeast, these proteins interact, the AD and BD are brought together, a functional transcription factor complex is constituted, and transcription of the downstream reporter gene is initiated which results in a visible phenotype in the yeast.

The interaction of the two B function MADS box transcription factors in *A. thaliana*, PI and AP3, has been proven *in planta* (McGonigle *et al.* 1996), however no interaction is seen in yeast (Yang *et al.* 2003) unless SEP3 is present (Immink *et al.* 2009). Before any experiments involving the introduction of the *P. vulgaris* MADS box proteins GLO and GLO^T into *A. thaliana* were attempted, the interactions of the *P. vulgaris* and *A. thaliana* MADS box proteins were tested in yeast. The assembly of the constructs was undertaken as part of this thesis, and the interaction experiments were performed by Dr. Barry Causier at the University of Leeds. The results from these analyses could help in understanding any phenotypes seen in transgenic *A. thaliana*.

The full-length sequences of *PI* and *AP3* were amplified by PCR and cloned into the pCR8 entry vector by way of TA cloning. The orientation and sequence were checked by single read sequencing, and the verified sequences were transferred into the destination vectors by way of topoisomerase I. The AD vector used was pGADT7 Rec,

and the BD vector was pGBKT7. Full details of this process are available in section 2.8.8, and plasmid maps for both vectors are included in the appendix.

The sequences of *PI* and *AP3* within the AD and BD vectors were checked by single read sequencing before use. AD and BD vectors containing *GLO*, *GLO^T* and *DEF*, the *P. vulgaris* orthologue of *AP3*, had already been made by Dr. Jinhong Li (Li *et al.* 2008).

Before the protein-protein interactions could be verified, it was first necessary to ensure that the yeast cells being tested contained both the AD and BD vectors. This was conducted to ensure that any lack of yeast growth was due to a genuine lack of interaction, rather than a problem with the yeast. Yeast cells containing the AD vector are capable of growth on media lacking leucine, and those containing the BD vector are able to grow on media lacking tryptophan, as a result, media lacking in both leucine and tryptophan is used to select for cells containing both. As shown in figure 5.6 panel (a), all cells are seen to contain both the AD and BD vectors.

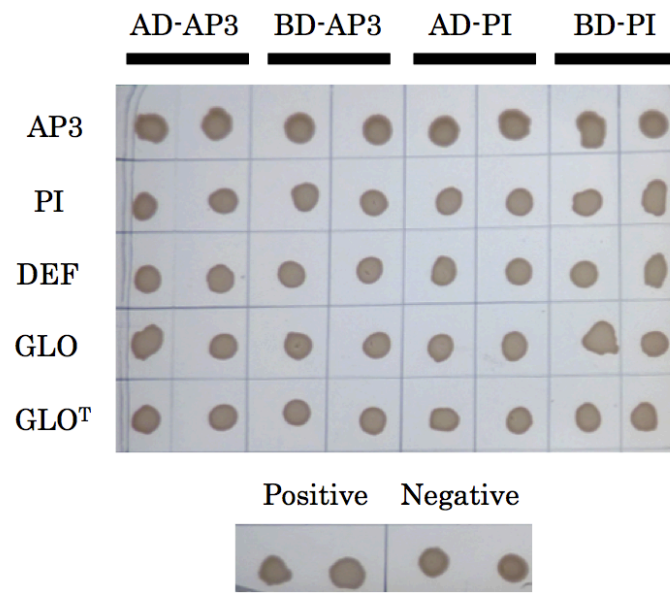
To test for interactions the reporter gene used in this experiment is *HIS3*; when a functional transcription factor is constituted by the interaction of an AD- and BD-fusion the *HIS3* gene allowed the growth of these yeast cells on media lacking histidine. In figure 5.5 panel (b), cells are grown on media lacking in leucine, tryptophan and histidine. Cells that are able to grow under these conditions must contain the AD and BD vectors, and a protein-protein interaction between the AD and BD.

No interaction is seen between AP3 and PI in any combination, however this is in agreement with previous data on the subject; the full-length *A. thaliana* proteins have previously exhibited a lack of interaction when tested in yeast (Yang *et al.* 2003), unless in the presence of SEP3 (Immink *et al.* 2009). The GLO-AD fusion shows an interaction with AP3-BD, however when in the GLO-BD does not show interaction with AP3-AD. In *P. vulgaris*, GLO dimerises with DEF (Li *et al.* 2008) to bring about the B function of the ABC model, so it is unsurprising that it interacts with the *A. thaliana* orthologue of DEF.

No interactions are seen between GLO and PI in any combination of AD and BD protein fusions. DEF and the *A. thaliana* orthologue of its partner, PI, show interactions in both combinations of AD and BD vectors, whereas GLO^T, unlike its paralogue GLO, does not interact with any of the *A. thaliana* proteins in any combination of AD and BD vector. As mentioned previously, Y2H experiments do

not always represent the interactions that are occurring *in planta*, however, the fact that there is a difference seen between the Y2H interactions of GLO and GLO^T may suggest that their functions have differentiated since the duplication event.

(a)



(b)

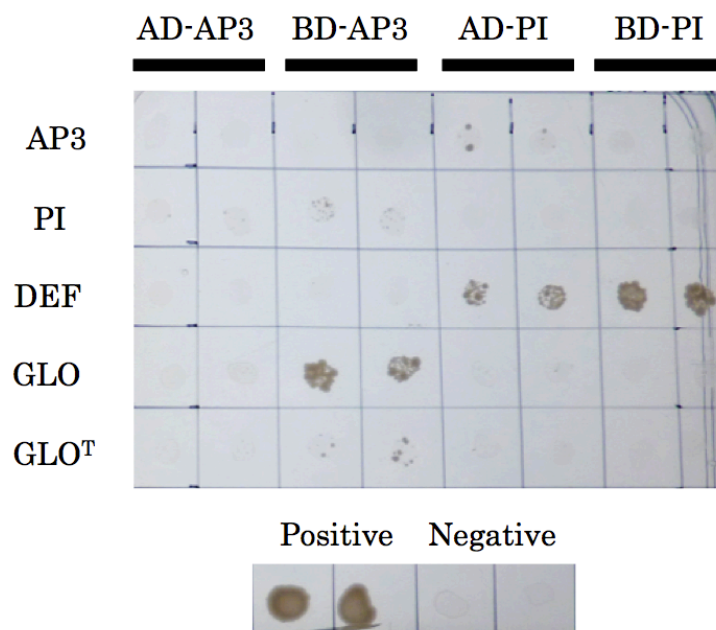


Figure 5.6: Yeast 2-Hybrid data for the AP3 and PI AD- and BD- fusions. Panel (a) shows yeast cells growing on media lacking in both leucine (selecting for AD vector) and tryptophan (selecting for DBD vector). Panel (b) shows yeast cells growing on media lacking in leucine (selecting for AD vector), tryptophan (selecting for DBD vector) and histidine (selecting for protein-protein interactions). The 'Positive' examples in both panels contain a yeast strain in which the AD- and BD-fusion are known to interact, and the 'Negative' a strain in which the AD- and DB-fusion do not interact.

5.5 Complementation of the *A. thaliana* *pi-1* mutant with *P. vulgaris* genes

The *pi-1* mutant displays the typical phenotype of a B function mutant; the transformation of petals to stamen, and of stamen to carpels. In the mutant, the function of the *PI* gene is altered by the presence of Single-Nucleotide Polymorphism (SNP), which introduces a premature stop codon (Goto and Meyerowitz, 1994). It has been shown that complementation of *pi-1* with the wild type *A. thaliana* *PI* restores the normal flower phenotype when driven by the native *PI* promoter. (Wuest *et al.* 2012).

PI shows a fairly high level of sequence similarity with its *P. vulgaris* orthologue *GLO*; 74% at the nucleotide level, with the 3' end of the coding sequence showing the highest number of differences. This, combined with the Y2H data, which have shown that the *P. vulgaris* proteins *GLO* and *DEF* are able, to some extent, to interact with the *A. thaliana* orthologues of their partners, demonstrates that their sequence is similar enough for the *P. vulgaris* to complement the *A. thaliana* gene. Studies in which the *PI* orthologue from other species has been able to complement the mutant have been successfully conducted; even in *Pisum sativum*, in which the floral structure is markedly different from that of *A. thaliana*, and the gene sequences show high levels of divergence, full complementation can still be achieved (Berbel *et al.* 2005). As a result, it is feasible that the *P. vulgaris* gene would be able to rescue the *pi-1* mutant, returning it to the normal flower morphology.

It is unknown whether *GLO^T* will be able to complement the *pi-1* mutant; its function in *P. vulgaris* is as yet unknown. As a MADS box transcription factor it will work as a dimer, and Y2H experiments have shown it to weakly interact with *DEF*, the partner of *GLO*. The interactions were not as strong as those seen between *GLO* and *DEF* (unpublished data, collected by Dr Barry Causier at the University of Leeds). If the function of *GLO^T* has diverged since the duplication event that created it, it will not complement the *pi-1* mutant. If *GLO^T* does complement the mutant, this would provide evidence that it is not the A function in the GPA model.

5.5.1 Construct design for complementation of the *pi-1* mutant

While the coding sequences of *GLO* and *PI* are constrained by their function and exhibit minimal divergence, the different floral architectures of *A. thaliana* and *P. vulgaris* may require different promoter regulation to bring about the correct expression patterns of these genes. In late stage flower buds, when the native *A. thaliana* *PI* promoter is driving expression of GUS, GUS can be detected in the second and third floral whorls (Honma and Goto 2000). When the promoter of the *Brassica napus* orthologue of *PI* is used to drive GUS expression in *A. thaliana*, GUS is can be detected in all four whorls in late stage flower buds (Roh *et al.* 2014).

For complementation to be possible, it was deemed important that expression of *GLO* and *GLO^T* were directed to the correct floral whorls of the flower bud, and it is possible that *P. vulgaris* promoter would not achieve this. As a result, both *P. vulgaris* genes were driven by the *A. thaliana* *PI* promoter, ensuring expression was directed to the second and third floral whorls.

The coding regions of *GLO* and *GLO^T*; were introduced into *A. thaliana*, driven by 2 kb upstream of the *PI* start codon and terminated by their own native terminators. Coding sequence only will be used as both *GLO* and *GLO^T* genomic DNA sequences contain very large introns. In addition, the coding region of *PI*, without introns, was also introduced into the *pi-1* mutant, under the control of its own promoter and terminator to ensure that the use of only the coding sequence does not impact on the complementation.

5.5.2 Construct production for the complementation of the *pi-1* mutant

The same cloning method will be used as described in section 5.2.1, as the *PI* promoter fragments designed for those experiments will be used in these constructs. The sequences of the genes for which the cDNA is required are already known and documented.

Figure 5.7 shows the constituent parts of each of the constructs in the order they will be arranged in the final binary vector. When a fragment has been split into two constituent parts, this indicates that a *BsaI* restriction site was present in the sequence and has been removed. *BsaI* sites present in coding regions were removed by the introduction of a silent mutation; the replacement codon was chosen on the

basis that it had previously been found to occur frequently in dicot species (Murray *et al.* 1989).

All fragments were amplified by PCR, using primers which added BsaI sites to the 5' and 3' ends of the amplified product. These products were cloned into pGEM-T Easy and sequenced by single-read sequencing. When digested, the overhangs left by the BsaI sites allowed the fragments to be ligated into the binary vector, which confers kanamycin antibiotic resistance in plants, in the correct order; the plasmid maps of the completed binary vectors containing all of the fragments are shown in the appendix.

The binary vector was checked as described in 5.2.2. Figure 5.8 (a) shows an example of the PCR reaction performed to confirm the correct ligation of constituent parts the binary vector after its introduction into *E. coli* DH5 α . Figure 5.8 (b) provides an example of the further verification of the completed binary vector; each join was sequenced via single read sequencing. Figure 5.7 (c) shows that the joins are still intact and all aspects are still present following introduction of the binary vector into *A. tumefaciens*. These constructs were introduced into *A. thaliana* Ler plants via the floral dip method, as described in section 2.9. The *A. tumefaciens* strain for *A. thaliana* transformations of the complementation vectors was GV3101 (section 2.8.9.5).

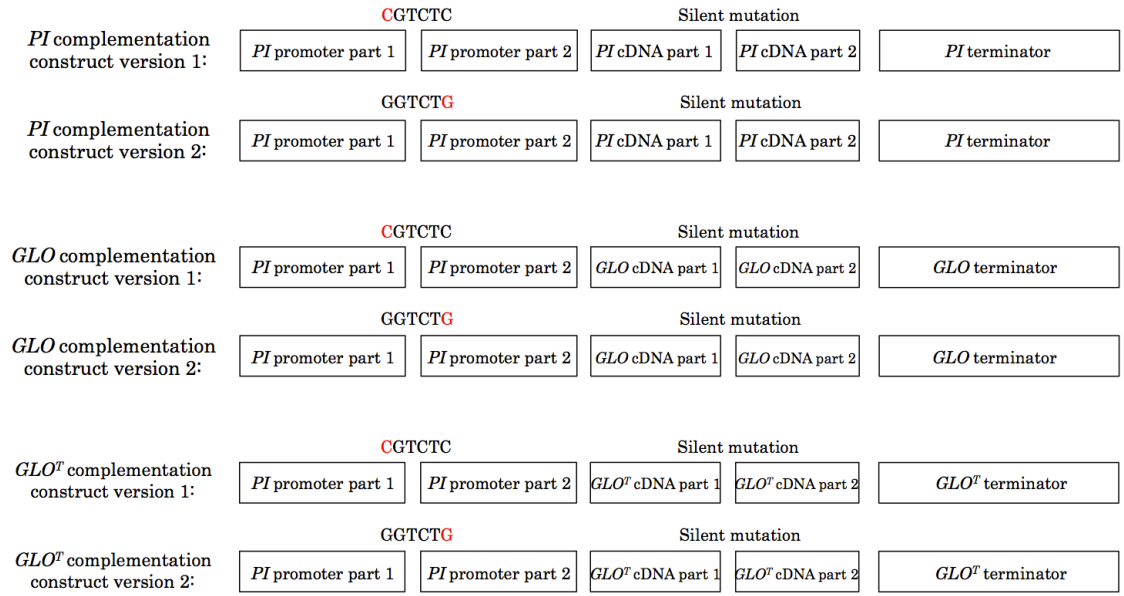


Figure 5.7: The six complementation constructs. Constituent parts of the six complementation constructs in the order they will be arranged when transferred to the binary vector. The two versions of the *PI* promoter are indicated; sequences above the boxes indicates the BsaI recognition site, and the base that has been changed to remove the BsaI site is highlighted in red. For each construct it is indicated where the cDNA sequence has been split into two parts to remove a BsaI site with the introduction of a silent mutation.

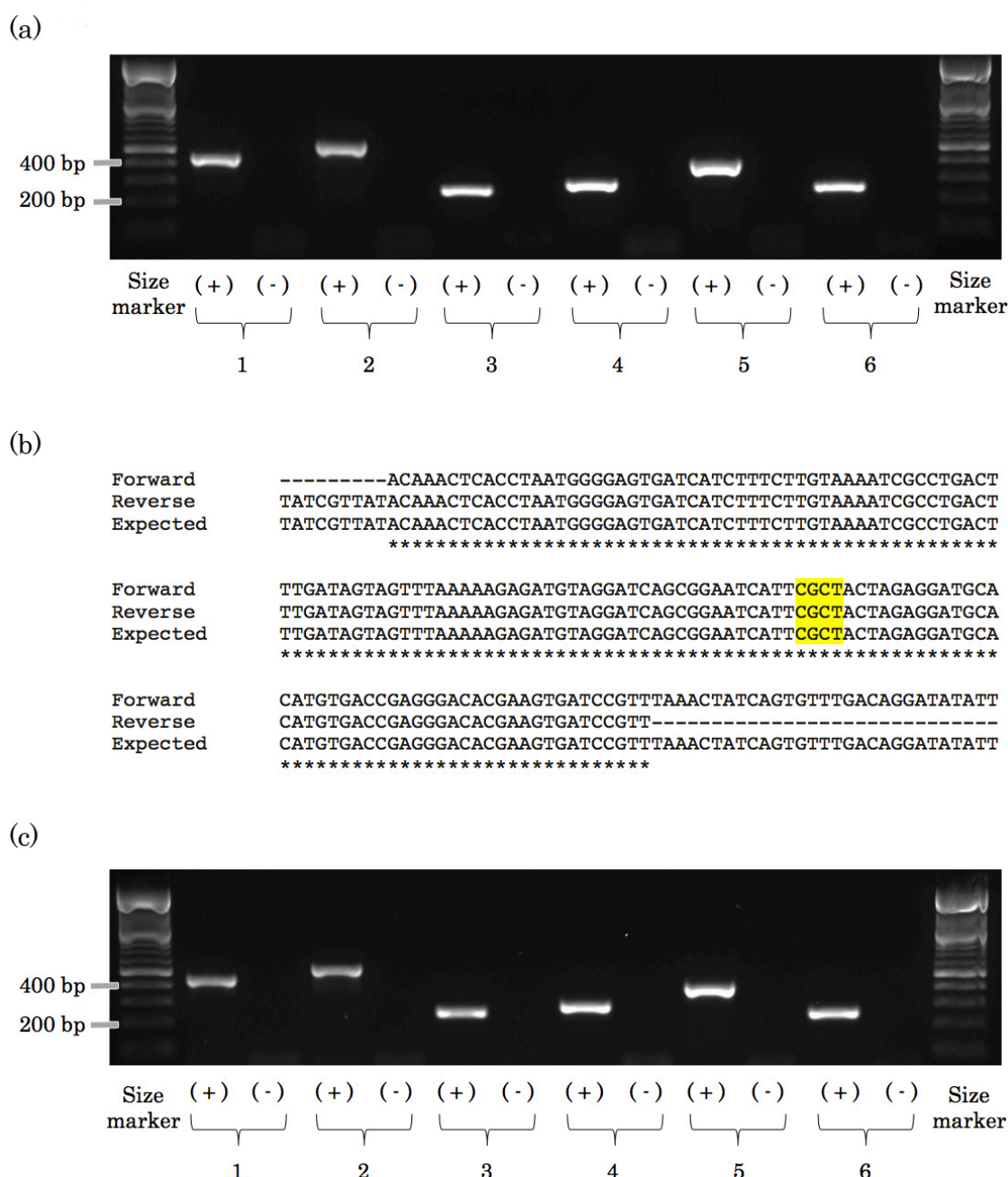


Figure 5.8: Example of the analyses performed on the complementation constructs.

(a) Standard GoTaq PCR conducted on *GLO* complementation construct 1 plasmid DNA using primers designed to verify that constituent parts of the construct have ligated in the correct order. Joins being tested are: 1 = binary vector and *PI* promoter part 1, 2 = *PI* promoter part 1 and *PI* promoter part 2, 3 = *PI* promoter part 2 and *GLO* coding region part 1, 4 = *GLO* coding region part 1 and *GLO* coding region part 2, 5 = *GLO* coding region part 2 and *GLO* terminator, 6 = *GLO* terminator and binary vector. (+) indicates plasmid DNA, (-) indicates a no DNA control. (b) DNA sequence alignment of two joined fragments. Expected sequence was compared to sequencing data obtained in the forward and reverse orientations. The example shown is combination 6 from panel (a); the joined overhang is highlighted in yellow. (c) Standard GoTaq PCR conducted on *A. tumefaciens* containing the *GLO* complementation construct 1, using the primer combinations from panel (a).

5.5.3 Analysis of *pi-1* complementation transformants

Due to time limitations the potential transformants that resulted from this investigation were not fully investigated. The plants transformed in this investigation were heterozygotes; the *pi-1* homozygous plants are deficient of anthers, and as a result these plants are unable to self-pollinate. Due to transforming heterozygotes, the likelihood of a transformant being a *pi-1/pi-1* homozygote was low, as only 25% of the offspring of the transformants would display this phenotype, and the transformation efficiency of Ler is lower than that of Col-0 (Clough and Bent 1998). The results of this experiment, and the potential for its continuation in the future are discussed in Chapter 6.

5.6 Overexpression of *GLO* in *A. thaliana*

In addition to attempting to complement the *pi-1* mutant in order to determine whether *GLO^T* specifies floral organ identity like the B function gene it is duplicated from, *GLO*, overexpressing each of these genes and analysing the transformant phenotype for both genes was proposed to yield information into the functionality of *GLO^T*. If the phenotype of overexpressing both genes was similar, it could be argued that both genes are performing the same or a similar function.

Work conducted by Dr. Sadiye Hayta has shown that the overexpression of *GLO^T* differs from that of the native *PI* overexpression; *GLO^T* overexpression plants show more similarity to the *cauliflower* mutant (Bowman *et al.* 1993) than to that of *PI* overexpression, in which there is a partial conversion of sepals to petal (Krizek and Meyerowitz 1996). Analysing the effect of *GLO* overexpression in *A. thaliana* will provide a point of comparison.

5.6.1 Construct production for the overexpression of *GLO*

The overexpression of *GLO^T* was conducted by Dr. Sadiye Hayta using the pBRAC^T 114 overexpression vector, which was obtained from Mr. Mark Smedley (Smedley and Harwood 2015). The pBRAC^T vectors operate using a dual binary system, in which pBRAC^T is always co-transformed with pSOUP. Sequences chosen for overexpression are initially TA cloned into the entry vector pCR8, and are transferred into the pBRAC^T vector by way of topoisomerase I activity; the full details of this are available in methods section 2.8.8. The full coding region of *GLO*

was used for the overexpression vector, as was the case with *GLO^T*, and the primers used to amplify this region are listed in the appendix.

The sequence of the *GLO* coding region was verified to contain no errors by way of single read sequencing, which was performed when the sequence had been cloned into pCR8; this is shown in figure 5.9 (a). The presence of the sequence was confirmed after its transferral into the pBRACT overexpression vector by way of PCR; this is shown in figure 5.9 (b). After the co-transformation of the overexpression vector and pSOUP into *A. tumefaciens*, a PCR reaction was conducted to ensure that the *GLO* sequence was still present; this is shown in figure 5.9 (c).

The *A. tumefaciens* strain for *A. thaliana* transformations of the pBRACT 114 and pSOUP vectors was AGL1 (section 2.8.9.6), and the Col-0 ecotype was used for these transformations, as this was used in the *GLO^T* overexpression transformation experiments. Full details of the floral dip procedure are available in section 2.9.

(a)

Reverse	TTTTTTATAATGCCAACTTTGTACAAAAAGCAGGCTCCGAATTCGCCCTT	<u>AGAGATGGG</u>
Forward	TTTTTTATAATGCCAACTTTGTACAAAAAGCAGGCTCCGAATTCGCCCTT	<u>AGAGATGGG</u>
Expected	-----	<u>AGAGATGGG</u>

Reverse	<u>TAGAGGAAAAATAG</u> AGATAAAGAGGATTGAGAATCAAAATAGGCAAGTTACTTATTC	
Forward	<u>TAGAGGAAAAATAG</u> AGATAAAGAGGATTGAGAATCAAAATAGGCAAGTTACTTATTC	
Expected	<u>TAGAGGAAAAATAG</u> AGATAAAGAGGATTGAGAATCAAAATAGGCAAGTTACTTATTC	

Reverse	AAAGAGGAGAAATGGGATCATAAAAAAGGCAAGGAGATCTCAGTTTATGTGATGCTCA	
Forward	AAAGAGGAGAAATGGGATCATAAAAAAGGCAAGGAGATCTCAGTTTATGTGATGCTCA	
Expected	AAAGAGGAGAAATGGGATCATAAAAAAGGCAAGGAGATCTCAGTTTATGTGATGCTCA	

Reverse	GGTCTCCCTTGTTATTTTGGCAACTCTGGTAAATGCATGAATATGCAGCCCTAAAC	
Forward	GGTCTCCCTTGTTATTTTGGCAACTCTGGTAAATGCATGAATATGCAGCCCTAAAC	
Expected	GGTCTCCCTTGTTATTTTGGCAACTCTGGTAAATGCATGAATATGCAGCCCTAAAC	

Reverse	TCCGTTGATTAACATCTTGGATGCATACCAGAAGCAATCTGGGAACAGGTTGTGGGATGC	
Forward	TCCGTTGATTAACATCTTGGATGCATACCAGAAGCAATCTGGGAACAGGTTGTGGGATGC	
Expected	TCCGTTGATTAACATCTTGGATGCATACCAGAAGCAATCTGGGAACAGGTTGTGGGATGC	

Reverse	TAAGCATGAGAACCTCAGCAACGAAATAGAAAGGGTCAAGAAAGAGAATGATAATATGCA	
Forward	TAAGCATGAGAACCTCAGCAACGAAATAGAAAGGGTCAAGAAAGAGAATGATAATATGCA	
Expected	TAAGCATGAGAACCTCAGCAACGAAATAGAAAGGGTCAAGAAAGAGAATGATAATATGCA	

Reverse	AATTGAGCTCAGGCACCTTGAAAGGAGAAGATGTACAATCTTTGCACCACAAGGAGCTTAT	
Forward	AATTGAGCTCAGGCACCTTGAAAGGAGAAGATGTACAATCTTTGCACCACAAGGAGCTTAT	
Expected	AATTGAGCTCAGGCACCTTGAAAGGAGAAGATGTACAATCTTTGCACCACAAGGAGCTTAT	

Reverse	GTCCATTGAATCCGCCCTCGAAAAATGGACTTGCTTGTGTTCCGCAGAGAGAGATGGAGAT	
Forward	GTCCATTGAATCCGCCCTCGAAAAATGGACTTGCTTGTGTTCCGCAGAGAGAGATGGAGAT	
Expected	GTCCATTGAATCCGCCCTCGAAAAATGGACTTGCTTGTGTTCCGCAGAGAGAGATGGAGAT	

Reverse	TTACAGGATGGCAAGAGAAAAATTTGCTGACAAGGAAAGGGTACTGGAAGATGAAAACAG	
Forward	TTACAGGATGGCAAGAGAAAAATTTGCTGACAAGGAAAGGGTACTGGAAGATGAAAACAG	
Expected	TTACAGGATGGCAAGAGAAAAATTTGCTGACAAGGAAAGGGTACTGGAAGATGAAAACAG	

Reverse	GAGCCTTACTTACCAATGCACCACCTGGTGATGGATATAGAAGCGGGGAAATGGAATA	
Forward	GAGCCTTACTTACCAATGCACCACCTGGTGATGGATATAGAAGCGGGGAAATGGAATA	
Expected	GAGCCTTACTTACCAATGCACCACCTGGTGATGGATATAGAAGCGGGGAAATGGAATA	

Reverse	TGGATATAATTACAGTCTCAATGCCATTTTCCTTCCGGGTGCAACCGATTACGCCAAA	
Forward	TGGATATAATTACAGTCTCAATGCCATTTTCCTTCCGGGTGCAACCGATTACGCCAAA	
Expected	TGGATATAATTACAGTCTCAATGCCATTTTCCTTCCGGGTGCAACCGATTACGCCAAA	

Reverse	TTTACAGGAGAGGATTAAATATAAACTCAAAAAGGGCGAATTCGACCAGCTTTCTTG	
Forward	TTTACAGGAGAGGATTAAATATAAACTCAAAAAGGGCGAATTCGACCAGCTTTCTTG	
Expected	TTTACAGGAGAGGATTAAATATAAACTCAAAAAGGGCGAATTCGACCAGCTTTCTTG	

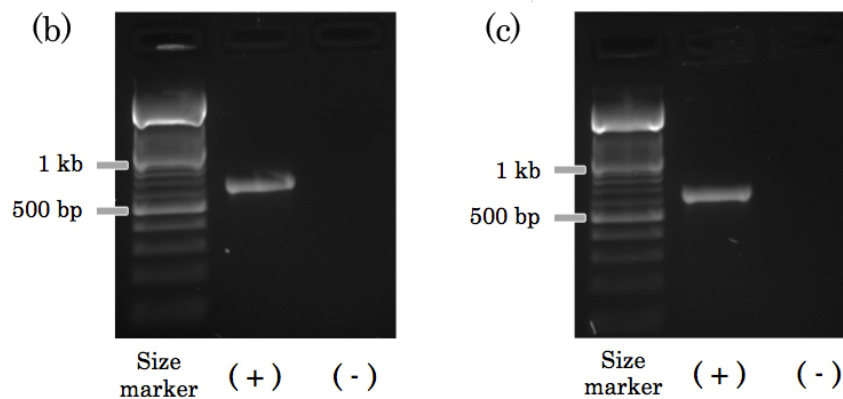


Figure 5.9: Analyses performed on the *GLO* overexpression construct. (a) Single read sequencing was performed on the pCR8 vector containing the *GLO* sequence. The forward primer used to amplify the *GLO* sequence is indicated by yellow highlight, reverse primer by green highlight. The start and stop codons are underlined. (b) Standard GoTaq PCR conducted on the overexpression vector to ensure the presence of *GLO* sequence, using the primers indicated in (a). The (+) indicates plasmid DNA, (-) indicates a no DNA control. (c) The presence of the *GLO* sequence was verified after the introduction into *A. tumefaciens*.

5.6.2 Analysis of *GLO* overexpression transformants

Due to time limitations, the transformants produced in this investigation were not fully analysed, and there was not time to generate the T₂ population required for adequate analysis of any floral mutations generated. The implications of this, and what the analysis of these mutants would bring to this investigation, are discussed in Chapter 6.

5.7 Summary of findings

Based on *in silico* analyses, the promoters of both *GLO* and *GLO^T* are proposed to contain numerous MADS box transcription factor binding sites, however it is only *GLO* that contains the binding sites for the AP3/PI dimer required for autoregulation of *PI* in *A. thaliana*. One of the AP3/PI transcription factor binding sites is found much closer to the start of the coding sequence than the other, and in *A. thaliana* there is a similar binding site that is required to maintain late-stage expression in *A. thaliana*. Based on the quantitative real time PCR data obtained in Chapter 3, the expression of *GLO* is maintained through flower development, and as a result this proposed regulatory element may be performing the same function as it is in *P. vulgaris*, however this would have to be experimentally verified.

Y2H data have shown different protein-protein interactions for *GLO* and *GLO^T*, when tested with the *A. thaliana* proteins AP3 and PI; *GLO* was able to interact with AP3 to some extent, however *GLO^T* was not. Although this system cannot accurately represent the interactions seen *in planta*, as shown by the lack of interaction between AP3 and PI, this further suggests that *GLO^T* may be performing a different function to that of *GLO*.

The findings of this chapter, and the experimental work which was unable to be completed during the timescale of this investigation, will be discussed in conjunction with data obtained in Chapters 2 and 3, where what is known thus far about the function of *GLO^T* will be considered.

CHAPTER 6

Discussion

6.1 Introduction

Heterostyly is defined as the existence of different flower forms within a species; *Primula vulgaris* exhibits a form of this, distyly. The two forms of *P. vulgaris* flowers, pin and thrum, exhibit reciprocal herkogamy, in which the sexual organs are positioned at alternate heights within each form (Webb and Lloyd 1986). In the thrum form the anthers are situated at the mouth of the corolla and the style is short, reaching approximately half way up the corolla tube. In the pin form the style is long, and is seen at the mouth of the corolla, whereas the anthers are positioned half way down the corolla tube. This difference in floral morphology has been suggested to increase out-crossing (Kohn and Barrett 1992; Lloyd and Webb 1992).

Finding the locus responsible for the different floral forms, the *S* locus, has long been of interest, and until recently the genetic components controlling the phenotypic differences seen between the *P. vulgaris* pin and thrum floral morphs remained undiscovered. Three functions have been defined as controlling the different arrangement of sexual organs that is seen between the pin and thrum flowers of *P. vulgaris*: *G*, *P* and *A* (Ernst 1936a). The *G* function is proposed to control the height of the style and the female aspect of the self-incompatibility system, the *P* function controls the size of the pollen and the male aspect of the self-incompatibility system, and the *A* function the height of the anthers (Ernst 1936a; Dowrick 1956).

Recent work has located a region of the *P. vulgaris* thrum genome that is absent from that of the pin (Li *et al.* 2016). This hemizygous region contains five thrum-

specific genes; one of these genes, *GLO^T*, forms the main focus of this study. In short homostyle *P. vulgaris* plants, in which both the style and the anthers are situated at approximately half way down the corolla tube, the hemizygous region of the genome is present, however *GLO^T* is not expressed (Li *et al.* 2016). This lack of gene expression leading to the altered floral phenotype of the short homostyle has resulted in *GLO^T* being presented as a candidate for the gene controlling the height of the anthers (Li *et al.* 2016), as described by Ernst (1936b).

6.1.1 The duplication of *GLO*

A gene duplication event, occurring approximately 52.7 MYA, resulted in the generation of a thrum-specific copy of the B function floral development gene, *GLOBOSA (GLO)* (Li *et al.* 2016). At the start of this investigation it was not known that *GLO^T* was a separate gene, and was presumed to be a thrum-specific allele of *GLO* (Li *et al.* 2008). It was only when this gene was found to be spatially separated from *GLO* that it was realised that *GLO^T* was a novel gene (Li *et al.* 2016). Preliminary semi-quantitative polymerase chain reaction (PCR) experiments performed by Dr. Jinhong Li confirmed that *GLO^T* was expressed in only flower bud tissues, and specifically in the second and third floral whorls. This expression pattern appeared to mirror its paralogue. These results were not fully quantitative, and the localisation was only conducted at one developmental stage, 15 mm flower buds. As a result, further work was required to ascertain when and where *GLO^T* was expressed, and as a result, both temporal expression through development, and localisation were investigated.

6.2 The selection of genes for normalisation of *GLO* and *GLO^T* expression

Before any studies could be conducted to assess the difference in temporal expression of *GLO* and *GLO^T* by way of quantitative real time PCR (qPCR), appropriate normalisation genes were first identified and verified. The practice of using a single normalisation gene can lead to high levels of error, and as a result the use of multiple reference genes is strongly recommended (Vandesompele *et al.* 2002). Not only is the inclusion of more than one normalisation gene important, but all genes must also be validated as showing stable expression before use.

The expression of *GLO* and *GLO^T* was measured across the different stages of flower bud development: from 2mm buds to 20 mm. Validated reference genes taken from studies conducted on similar tissues, or across a large range of different tissues in other species were used as the basis for selection of the original cohort of candidates to be tested.

The *P. vulgaris* orthologues of the normalisation genes selected from expression studies in *A. thaliana* and other species were identified, using the *P. vulgaris* genome sequence and putative gene models available. Primers were designed and tested for stability of expression using a number of different validation methods; analysis of the raw Cq values, ΔC_t , Normfinder, and the GeNorm aspect of the Biogazelle qbase+ software suite (Andersen *et al.* 2004; Silver *et al.* 2006; Hellemans *et al.* 2007). While there were variations between which genes were ranked as the most stably expressed, thus proving the importance of performing more than one test, one *P. vulgaris* gene, *ACT* was consistently ranked as the least stably expressed gene in all tests, and was excluded from use in normalisation of the data. As the remaining three genes, *ELF α* , *TUA* and *PP2A*, all performed well and were used for normalisation of gene expression studies. Following the successful validation of these *P. vulgaris* normalisation genes, the expression of *GLO* and *GLO^T* could be accurately analysed.

6.3 Temporal expression of *GLO* and *GLO^T* expression

The expression of *GLO* and *GLO^T* was measured across five flower bud developmental stages; 2 mm, 5 mm, 10 mm, 15 mm, and 20 mm. During this time, pin and thrum flowers undergo changes, which transform them from being identical at 2mm, to looking markedly different at 20 mm. These different stages were selected based on previous work on floral development (Webster and Gilmartin 2006), but with minor modifications due to differences in floral cultivar used between the experiments. *GLO* expression is seen to increase through development, increasing steadily from its lowest point at the 2 mm bud stage, to a peak at 20 mm buds.

Until these experiments were undertaken little was known about *GLO* expression through the later stages of flower development in *P. vulgaris*; it had been detected in floral meristems, and semi-quantitative PCR had shown expression in larger flower buds (Li *et al.* 2010), however its expression dynamics through the later floral stages were unknown. Research conducted on the orthologues of *GLO*, *GLOBOSA*

(*GLO*) in *Antirrhinum majus* and *PISTILLATA (PI)* in *A. thaliana*, have shown that not only are these genes expressed in both early and late stages of petal differentiation, but that they are present up until the point of anthesis (Honma and Goto 2000; Manchado-Rojo *et al.* 2012).

The continued expression of *GLO* to the point just before the opening of the flowers, 20 mm, appears to be in concurrence with the expression of its orthologues. This new information regarding the expression of *GLO* provided a standpoint to which the expression of *GLO^T* could be compared, and was key to answering the question as to whether the expression patterns of *GLO^T* had diverged from those of *GLO*.

The expression patterns of *GLO^T* vary from those of *GLO*. *GLO^T* is expressed at a lower level than *GLO* in all of the flower bud stages tested, and does not show the gradual increase through the stages of development. *GLO^T* expression is not significantly raised during the developmental stage in which the anthers begin to elevate in the thrum flowers, nor at any other stage. This would suggest that *GLO^T* expression may be regulating a gene or number of genes downstream of itself to bring about the raising of the anthers. While the downstream targets of *GLO* and *GLO^T* are not known, *PI* has been shown to regulate only a small number of genes, which do not display tissue specificity, and are potentially involved in other pathways (Zik and Irish 2003).

While these experiments have been able to provide information as to the difference in expression between *GLO* and *GLO^T*, they are unable to provide localisation within the flower bud. Quantitative expression analyses on the floral whorls would also only provide so much data, as can the non-quantitative localisation experiments that have been conducted (Chapter 4). Quantitative expression analyses of *GLO^T* which would be able to provide some localisation data could answer some of the remaining questions that remain about the expression of *GLO^T*, even if less precise than that of RNA *in situ* hybridization.

If *GLO^T* is controlling the anther elevation, it may only be expressed within a narrow band of tissue within the second and third whorls, and this may be a potential difference between the expression of *GLO* and *GLO^T*. Dissection of the petal tissue into different sections – corolla tube below point of anther attachment, corolla tube at anther attachment, corolla tube above point of anther attachment – was attempted during this investigation; however the time required to conduct this dissection resulted in RNA degradation and it was determined that this would bias results. However, further investigation into this is strongly suggested.

6.4 Localisation of *GLO* and *GLO^T*: RNA *in situ* hybridisation

The RNA *in situ* hybridisation of *GLO* had been previously shown, and expression was localised to the second and third floral whorls of *P. vulgaris* floral meristems (Cook 2002). These results were obtained before the existence of *GLO^T* was known, and as a result both pin and thrum floral meristems were included in experiments with no distinction being made between the two. Due to the high level of similarity in the sequence of *GLO* and *GLO^T*, it was decided that this experiment needed to be repeated to define the individual expression patterns of both genes.

The localisation of *GLO* showed the same expression pattern as in previous experiments on *P. vulgaris*, the second and third whorls of the developing floral meristems (Cook 2002), and this mirrors what is seen in the *A. thaliana* and *A. majus* orthologues (Tröbner *et al.* 1992; Goto and Meyerowitz 1994). Expression appears to be consistent between both pin and thrum samples, and the lack of different localisation patterns between the two suggests that either *GLO^T* localises to the exact same tissue types as *GLO* or that the probe used is specific enough for this not to cause issue.

For the localisation of *GLO^T*, the specificity of the probe is easier to test; due to the confirmed lack of expression of *GLO^T* in pin tissue, the inclusion of pin samples were able to act as an internal control for background staining and cross-hybridisation. Localisation of *GLO^T* does not appear to have the same boundaries of expression as that of *GLO*; it is not tightly constrained to the forming second and third whorls as *GLO* and its *A. thaliana* and *A. majus* orthologues. Levels of background staining are higher than those in the *GLO* localisation, however it is distinctly lower than that of the sense control slides and those of the pin antisense slides.

Further to the different expression patterns shown in the floral meristems, attempts were made to analyse the expression in larger flower buds; the same samples as tested under qPCR conditions were attempted: 2 mm, 5 mm, 10 mm, 15 mm, and 20 mm. However, even in the youngest tissue lignification is beginning to occur, and presents as brown colouration of the tissues that is present before *in situ* experiments have taken place. In addition to this, cells in these later stages of floral development were significantly expanded, and cell contents appeared to be being pushed against the cell wall, resulting in a ring of staining around the cell wall, which was impossible to distinguish from background levels.

The application of RNA *in situ* hybridisation for the later stages of floral development may be limited; numerous modifications were made to the protocol, however the localisation seen in the floral meristems could not be replicated even when the same probe was used on the larger bud tissue. As a result, other methods of localising expression in larger *P. vulgaris* must be attempted. The production of transgenic *P. vulgaris* into which promoter-reporter constructs have been transformed is currently underway, and this method could be used to bypass the problems found with RNA *in situ* hybridisation.

6.5 Localisation of *S* locus proteins: design of peptide antibodies

The design and synthesis of peptides that would result in antibody production that did not recognise paralogues of the protein of interest was determined to be of key importance in this study, particularly in the case of GLO and GLO^T. The results of RNA *in situ* hybridisation experiments suggest that the expression patterns of *GLO* and *GLO^T* differ, and the addition of localising the expression of the protein would further provide details as to the action function of GLO^T in comparison to GLO. Peptides were also designed to the protein sequences of two other genes in the *S* locus: *PUM^T* and *KFB^T*. *CYP^T* and *CCM^T* both displayed paralogues that were of a high enough level of sequence similarity that even a short peptide sequence could not be found that would show specificity and result in a functional antibody being raised to it.

While there was not enough time to assess the success of the antibodies produced, their design and production will provide valuable resources to discover more about the localisation of GLO and GLO^T, and also of PUM^T and KFB^T. Before use, all antibodies must be checked by Western blot, to first ensure that they will recognise the peptide to which they were raised. The subsequent stage of assessment would be to test the antibody via a Western Blot against a crude protein extract obtained from both pins and thrums flower buds, before immunolocalisation experiments could be conducted. Floral meristems are the recommended tissue to attempt preliminary experiments on, due to the success seen with RNA *in situ* hybridisation, and colour development methods of detection, for example the alkaline phosphatase NBT/BCIP method used in RNA *in situ* hybridisation experiments carried out as part of this study, are strongly suggested for use in these experiments due to the high levels of autofluorescence exhibited by *P. vulgaris* bud tissues, as determined by spectral analysis by Dr. Sadiye Hayta (data not shown).

6.6 Interactions of GLO and GLO^T with other proteins

In order to learn more about the functions of the GLO and GLO^T proteins, their interactions with their *A. thaliana* orthologues and the orthologues of their usual partners were assessed by way of Yeast 2-Hybrid experiments. Not only was the intention of these experiments to inform the design of constructs in which *P. vulgaris* sequence would be expressed in *A. thaliana*, but to determine whether GLO and GLO^T interacted with *A. thaliana* proteins differently.

The cDNA sequences of the *A. thaliana* B function MADS box genes *PI* and *APETALA3 (AP3)* were isolated and cloned into Activation Domain (AD) and DNA-Binding Domain (BD) vectors for expression in yeast (as shown in Chapter 5). Mating tests were conducted against the *A. thaliana* sequences and AD and BD-fusions of the cDNA sequences of *GLO*, *GLO^T* and *DEF*; the interaction experiments were carried out by Dr. Barry Causier at the University of Leeds. The only strong interactions seen in these experiments were between GLO in the AD-fusion and AP3 in the BD-fusion, and between DEF and PI in both combinations of AD and BD-fusions. No interactions were seen between PI and AP3 in either combination of AD and BD-fusions, however this is a result that has been reported previously (Yang *et al.* 2003). The fact that the *P. vulgaris* proteins are able to interact with the *A. thaliana* proteins in yeast suggests that their introduction into *A. thaliana* could result in successful complementation, although this is not guaranteed.

That GLO and GLO^T do not show the exact same interactions when paired with *A. thaliana* proteins would suggest there are differences in their binding partners, resulting in different downstream effects. These data are supported by the preliminary mating tests conducted in which the *P. vulgaris* proteins listed above were all tested against each other. This study has found that GLO^T shows a potential weak interaction with DEF the partner of its paralogue, GLO, however it is not known if this interaction is seen *in planta*, and will require further experiments. Efforts are currently in progress to identify further protein-protein interactions of GLO^T; Dr. Barry Causier is currently in the process of building cDNA library for *P. vulgaris*.

6.7 Promoter analysis of *GLO* and *GLO^T*

Since the duplication event, which brought about the creation of *GLO^T*, it has thus far been shown that there are differences between the two paralogues in terms of expression. In order to discover some of the potential causes of this, the promoters of the two genes were analysed. It has been shown that in *A. thaliana* auto-regulation by the AP3/PI heterodimer plays a part in the expression of *PI*, however it is not known to what extent the promoter of *GLO^T* has changed since the duplication. As a result, 2 kb upstream of the start codon was analysed for both genes, and the MADS box transcription factor binding sites were analysed.

The promoters of *GLO* and *GLO^T* both show a number of MADS box transcription factor binding sites; *GLO* contains 14, *GLO^T* contains 9. While these results are only a prediction of the binding sites of the MADS box proteins, and all of these may not be functional *in planta*, distinct differences are seen between the promoters of *GLO* and *GLO^T*. *GLO^T* lacks any predicted PI/AP3 binding regions in the promoter, whereas *GLO* contains two. This divergence of binding sites could explain the different localisation patterns seen between the two genes in RNA *in situ* hybridisation experiments, and the different levels of expression seen when analysed via qPCR.

6.8 Design of constructs for transformation of *A. thaliana* with *P. vulgaris* genes

A number of constructs were designed and produced for the transformation of *A. thaliana* with *P. vulgaris* promoter sequences and coding regions. Analysing the effects of these genes in a heterologous system, and how *GLO* and *GLO^T* interact with *A. thaliana* proteins would provide information as to the potential divergence of *GLO* and *GLO^T*. These experiments intended to determine whether the regulation of these genes has altered since the duplication of *GLO* and whether the proteins perform different functions.

6.8.1 Promoter-reporter constructs

Based on the promoter analysis for *GLO* and *GLO^T*, the promoter of *GLO* was found to have at least one site to which the AP3/PI dimer would bind, and hence, expression would be brought about. While a number of MADS box recognition sites were found in the *GLO^T* promoter, none were similar enough to the PI/AP3 site. In order to

determine whether the *GLO* promoter was similar enough to that of *PI* to initiate transcription of the downstream coding sequence, a promoter-reporter construct was designed, in which ~1.8 kB of the *GLO* and *GLO^T* promoter sequences were used to drive GFP, followed by a NOS terminator. As a control, the native *PI* promoter was also used to create a promoter-reporter construct.

A. thaliana plants were transformed with these constructs, however due to time constraints the transformants were unable to be analysed; the expression of these constructs will only be present in the floral tissues, and as a result analysis could not take place until plants had reached maturity, and these analyses are suggested as future work to continue this project. Based on the promoter analyses, it would be expected that the expression of *GLO* would be regulated by the same elements as that of *PI*, as the *A. thaliana* PI/AP3 heterodimer binding domain was located within its promoter, and as a result would show expression in the second and third whorls. The lack of AP3/PI binding site within the *GLO^T* promoter may suggest that it will not be localised to the second and third whorls, however this could also be achieved through the action of other elements.

6.8.2 Complementation of the *pi-1* mutant with *P. vulgaris* genes

Informed by the Y2H results obtained when the *P. vulgaris* proteins GLO, GLO^T and DEF were tested for interactions with the *A. thaliana* proteins PI and AP3, there was evidence that *GLO* would be able to complement the *pi-1* mutant. *GLO^T* did not show interactions with any of the *A. thaliana* proteins, rendering it less likely to complement, however, Y2H experiments cannot accurately predict what will happen *in planta* and as a result it was determined that attempting to complement the *pi-1* mutant with both *GLO* and *GLO^T* would provide valuable information as to their functions. There was not time to analyse the transformants generated in this experiment, and as a result this is suggested as part of the future work.

6.8.3 Overexpression of *GLO* in *A. thaliana*

The creation and subsequent transformation of *A. thaliana* an overexpression construct in which the *P. vulgaris* gene *GLO^T* is expressed under the control of the 35S Cauliflower Mosaic Virus promoter was conducted by Dr. Sadiye Hayta, and the resultant transformants displayed an abnormal phenotype, which appeared similar

to the *cauliflower* (Kempin *et al.* 1995) mutant seen in certain accessions of *A. thaliana*.

Were *GLO^T* to still have retained the B function capabilities of its paralogue, the overexpression of this gene would usually result in the conversion of the first whorl from sepals to petals. It is possible that this phenotype is the product of the new function *GLO^T* is now performing; having diverged since the duplication event, or that in a heterologous system it does not function in the same manner as in *P. vulgaris*. It is equally possible that the lack of a gene orthologue for *GLO^T* has resulted in its pairing with a different MADS box protein to form a heterodimer.

As a result, it was decided that the experiment performed on *GLO^T* was to be mirrored by the overexpression of *GLO*. If a similar phenotype was seen in the *GLO* overexpression plants, it would suggest that both *P. vulgaris* proteins do not show a high enough level of sequence similarity to the *A. thaliana* proteins to dimerise correctly and bring about their correct function, and as a result are acting erroneously. This is, however, entirely speculative at this point; due to time constraints the resultant progeny from these experiments were not analysed, and their analysis is suggested as further work.

6.9 Conclusions

From the experiments conducted as part of this study, a number of differences have been identified between *GLO* and *GLO^T*. Early in floral development, at which point the floral whorls are being formed, *GLO* and *GLO^T* appear to be showing different expression patterns; *GLO* expression is tightly defined to the second and third whorls, whereas *GLO^T* expression is more widespread. Later in development, the expression of *GLO* and *GLO^T* continues to differ, with *GLO* showing an increase in transcript levels when flower bud size increases, whereas *GLO^T* expression appears constant. Despite being a paralogue of *GLO*, *GLO^T* is showing different behaviour to *GLO* both early and late in development, and these differences may indicate that since the duplication event, *GLO^T* has undergone a change of function.

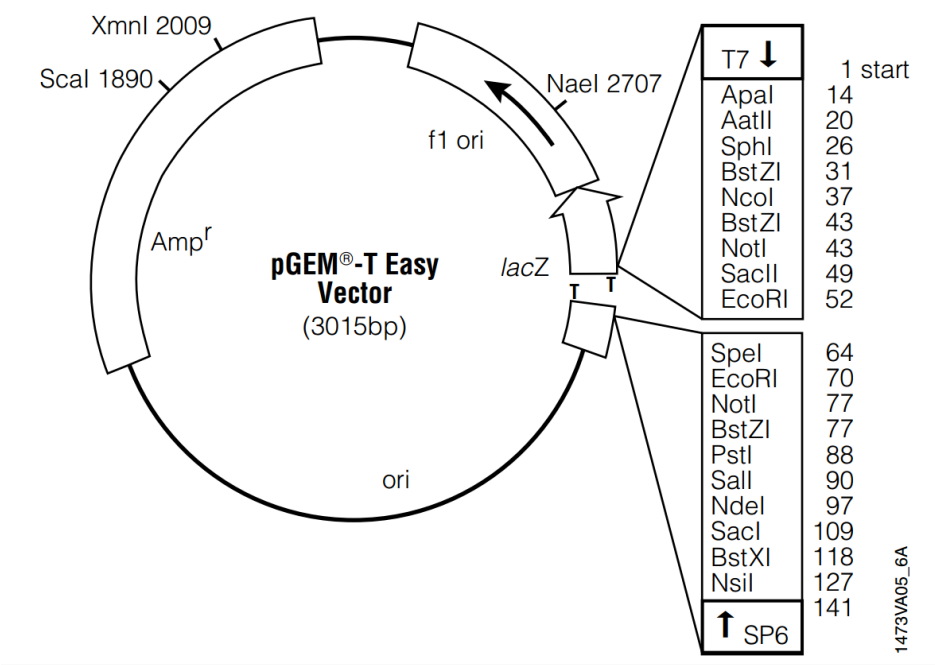
The promoter sequences of *GLO* and *GLO^T* contain different MADS box transcription factor binding sites, which may indicate different regulation, and the analysis of transgenic *A. thaliana* plants generated in this investigation will further expand knowledge as to the affect that these differences in promoter binding sites are bringing about, and the behaviour of the two different proteins in a homologous system. Further work focusing on the localisation of the *GLO* and *GLO^T* proteins will provide more evidence as to the potential difference in function seen between *GLO* and *GLO^T*.

GLO^T is the strongest candidate for the raised anthers seen in *P. vulgaris* that has thus far been identified, due to the combination of its localisation pattern and the short homostyle phenotype that is seen when it is not expressed (Li *et. al* 2016), however further work must be carried out before this can be verified. The fact that it shows different localisation and expression levels to its paralogue indicate that it is likely to be performing a different function.

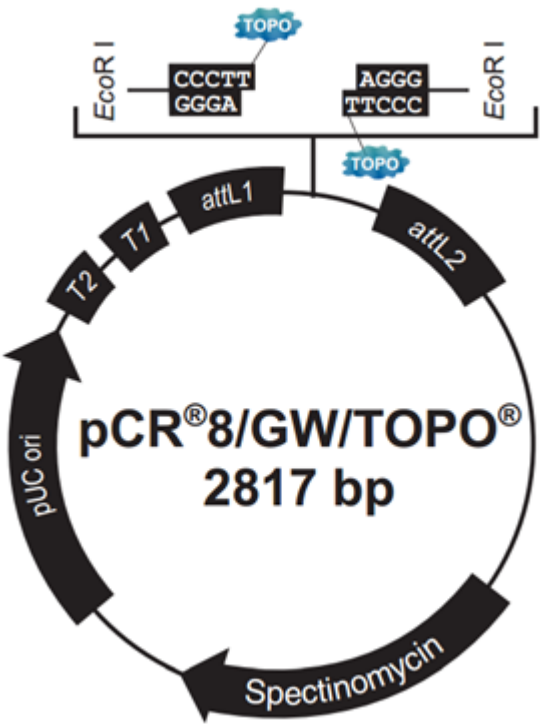
APPENDIX

Plasmid maps:

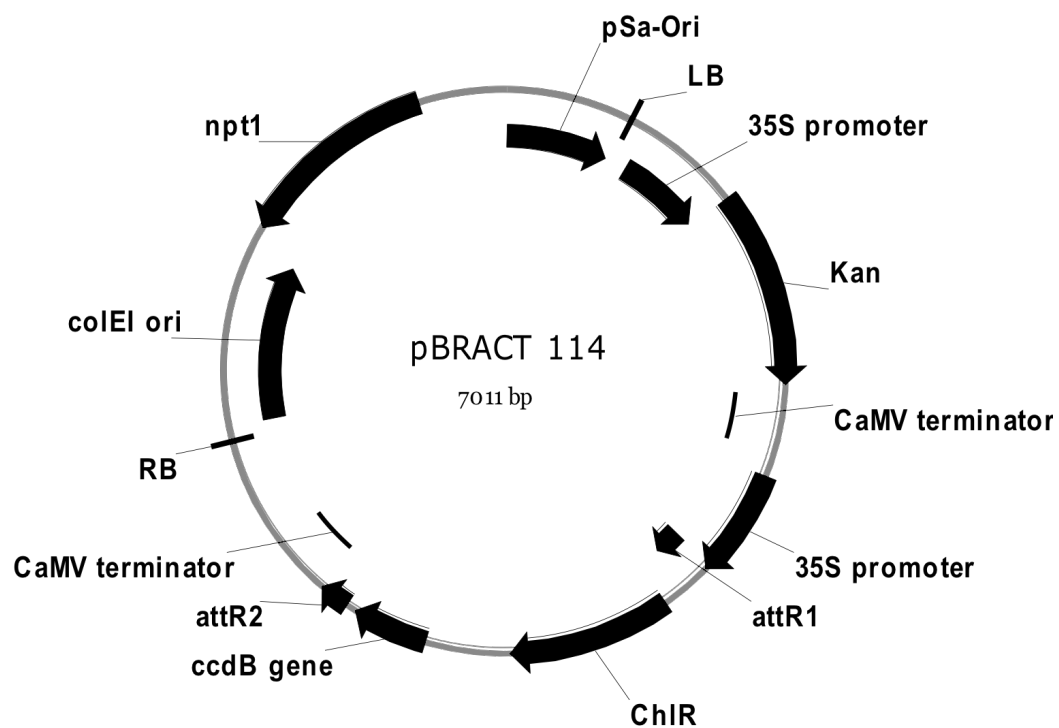
pGEM-T Easy (Promega UK Ltd.):



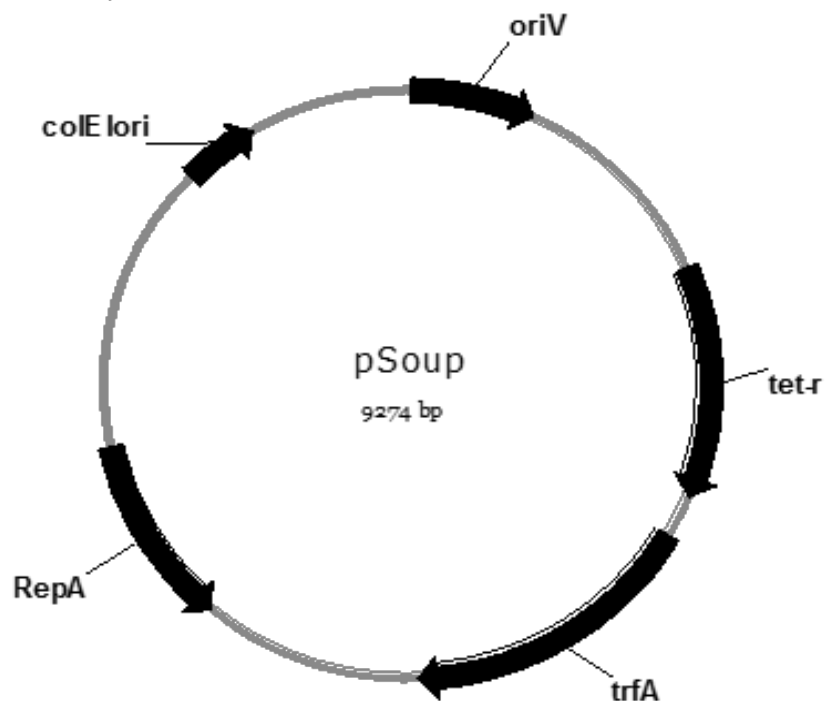
pCR8 (Thermo Fisher Scientific Ltd.):



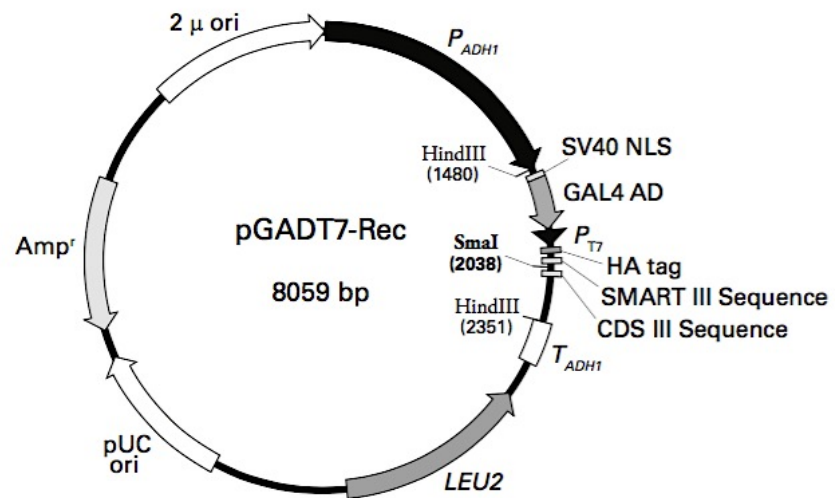
pBRACT 114 (Smedley and Harwood 2015):



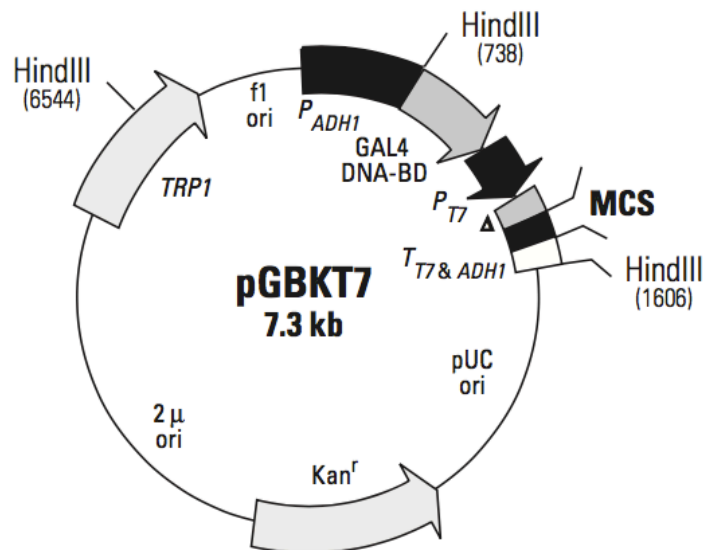
pSOUP (Smedley and Harwood 2015):



pGADT7-Rec (Takara Bio Europe)
Must be digested with *Sma* I before use:



pGBKT7 (Takara Bio Europe)



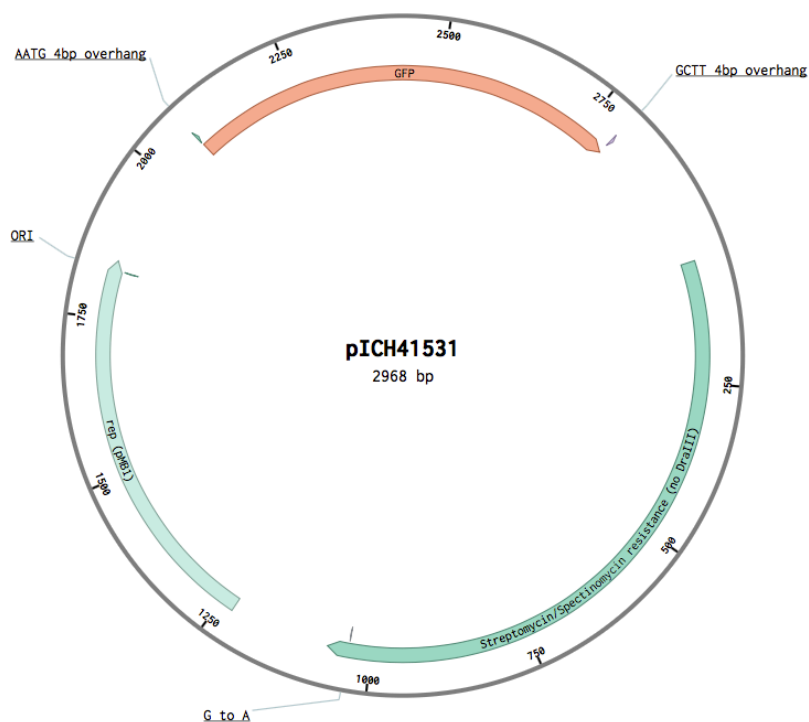
Plasmid maps of vectors from The Sainsbury Laboratory:

Binary vector: (GGAG and CGCT overhangs created when cut with BsaI)

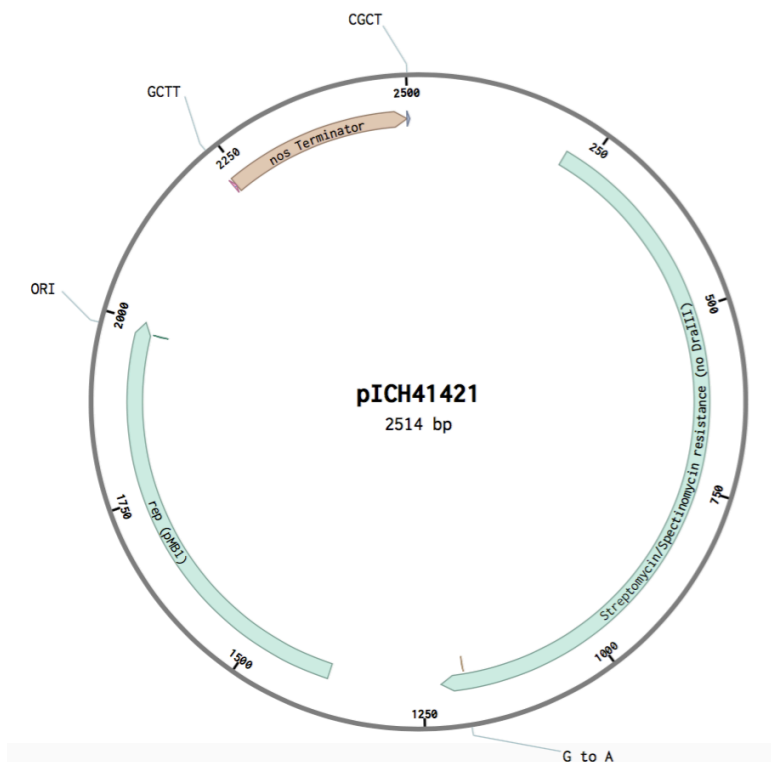
Plasmid map generated by SnapGene®



pICH41531 GFP coding region (Engler *et al.* 2014)
AATG and GGTC overhangs created when cut with BsaI
Plasmid map generated by Benchling:

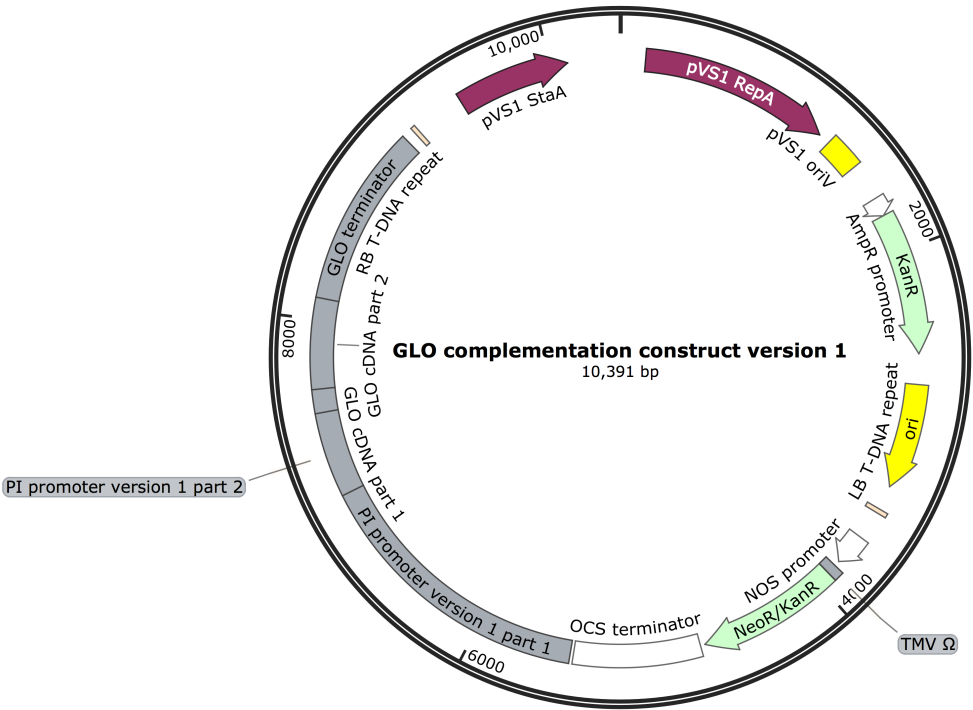


pICH41421 NOS terminator region (Engler *et al.* 2014)
GCTT and CGCT and the overhangs created when cut with BsaI
Plasmid map generated by Benchling:

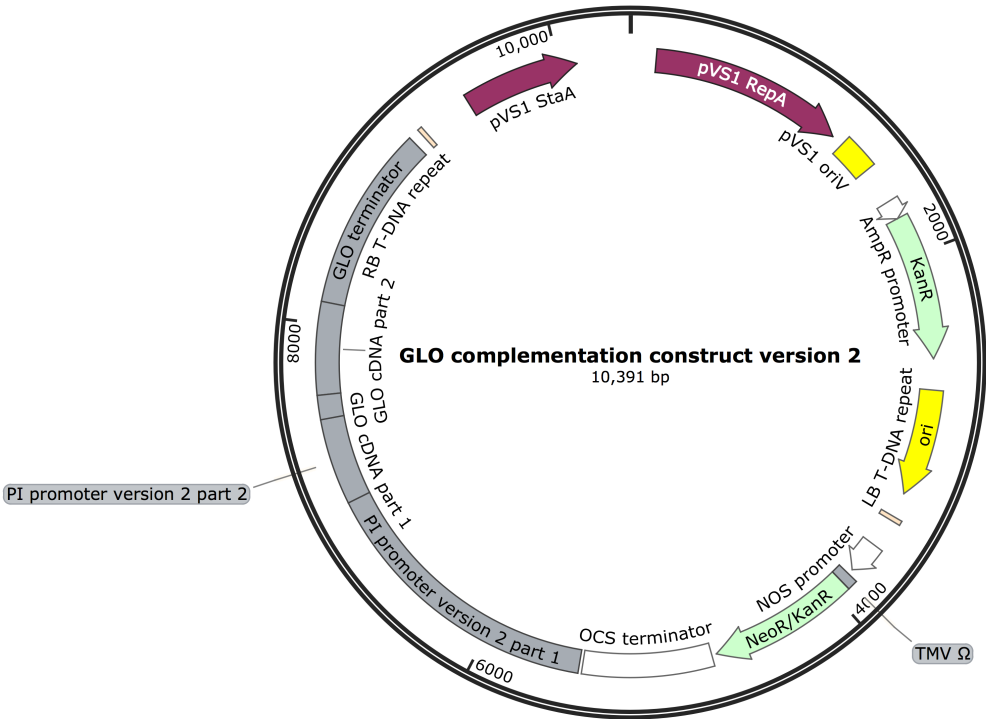


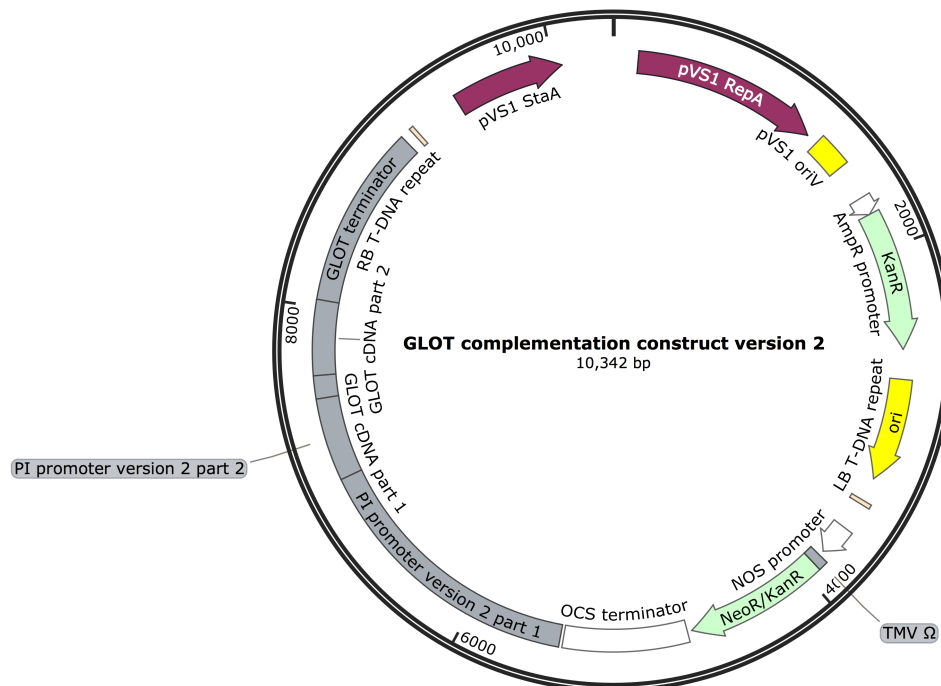
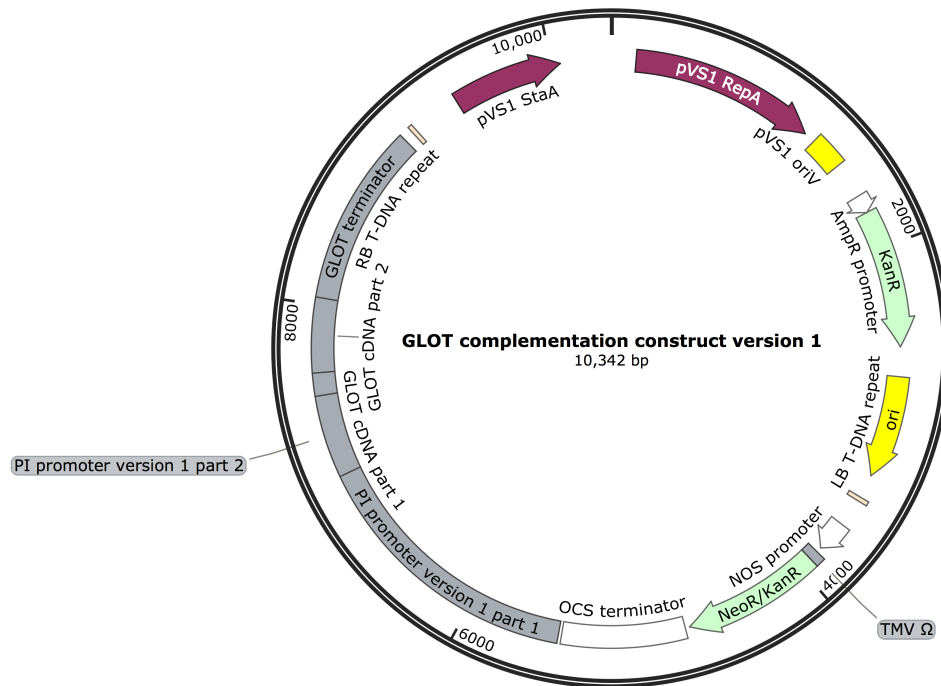
Plasmid maps of user-generated constructs:

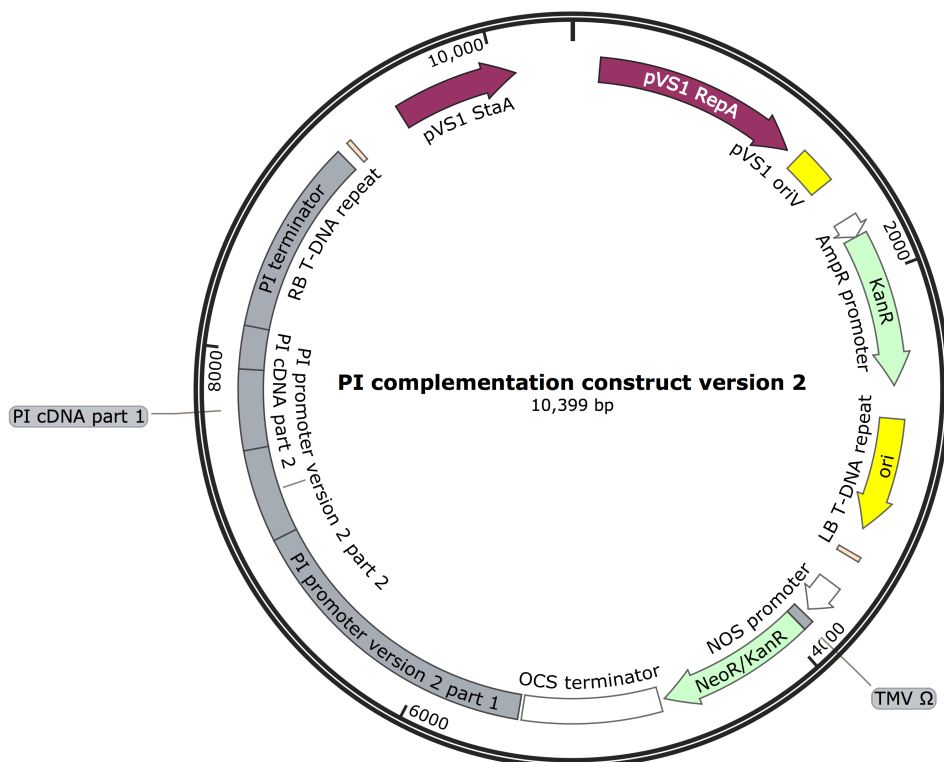
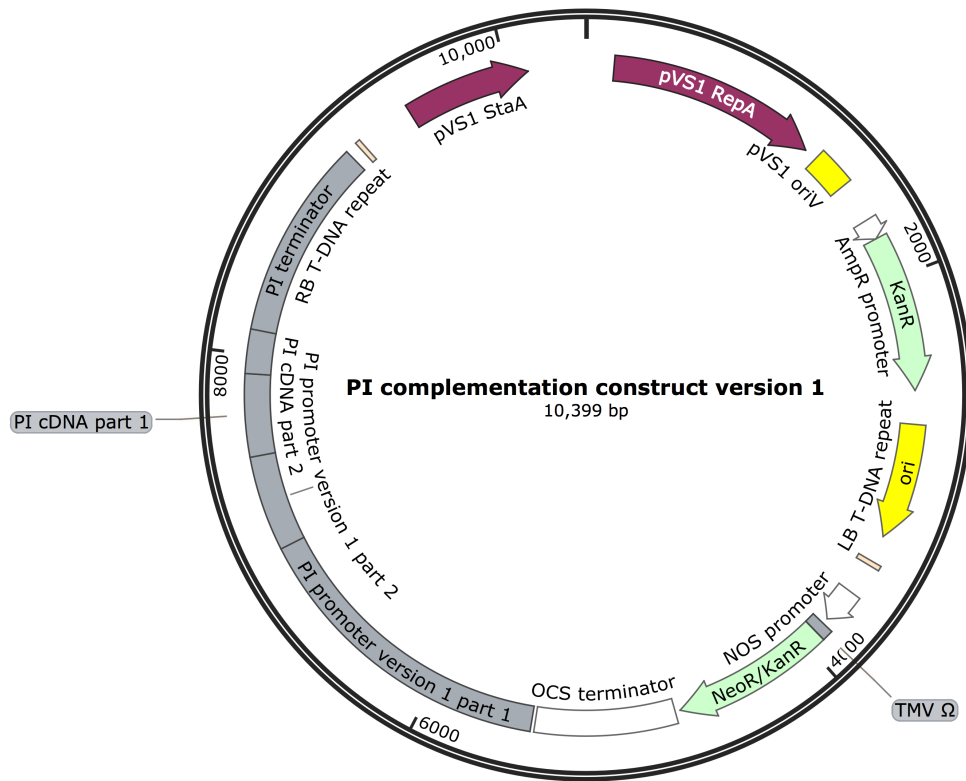
Created with SnapGene®

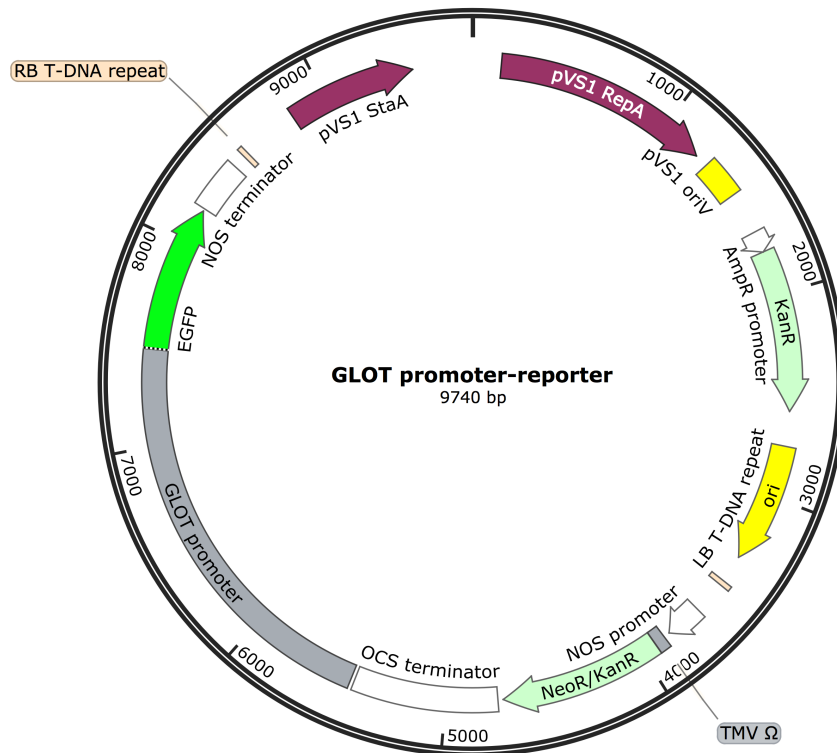
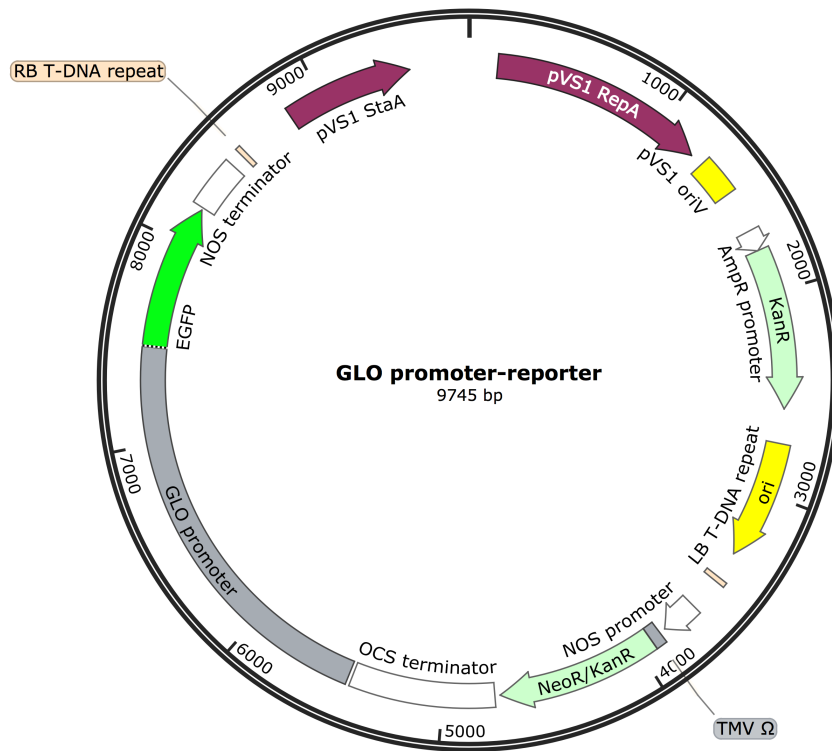


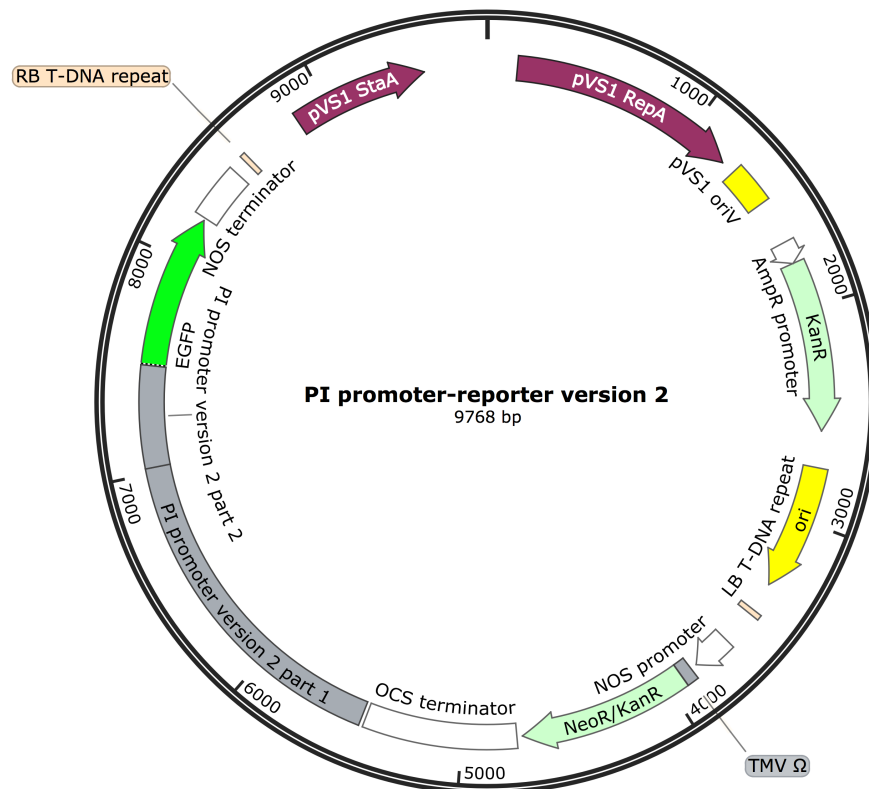
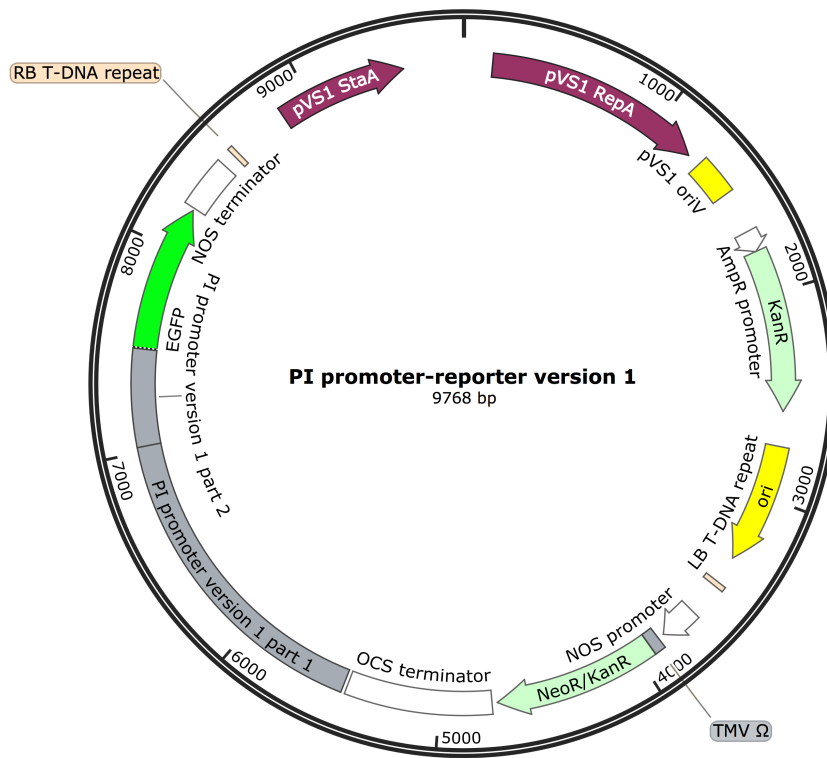
Created with SnapGene®











Primer sequences:

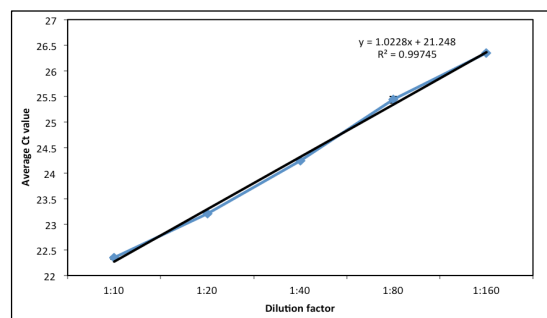
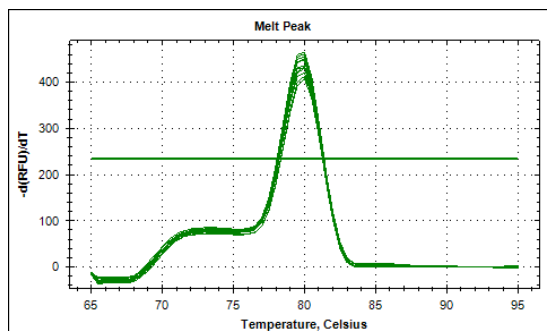
Primer name	Primer sequence 5'- 3'	Description
GLO F	AGT TAG CTA GAG AGG AAG AG	Isolation of cDNA sequence of <i>GLO</i>
GLO R	GTA GAG CAT TAC GAC AAT TTG	
GLOT F	GAG AAC AAG AAA GCT AGA GAG	Isolation of cDNA sequence of <i>GLOT</i>
GLOT R	CTC AAG GTT GAT ATT TCA GGT T	
ELF1 α 1F	TTA TCG ACT CGA CTA CTG GAG G	qPCR normalisation gene primer pair
ELF1 α 1R	GGT AGC GTC CAT CTT GTT ACA G	
PP2A 1F	TCA TGG GTG ACT ATG TTG ATC G	qPCR normalisation gene primer pair
PP2A 1R	ATT TGC CGA CTT TCG TGA TTC C	
PP2A 2F	GTC AGG TTT TCT GTT TGC ATG G	qPCR normalisation gene primer pair
PP2A 2R	CAA AAG TAT ATC CAG CTC CAC G	
PP2A 3F	GTC TCA GGA GAA GAA TGT GG	qPCR normalisation gene primer pair
PP2A 3R	TCT GGC TCA ATT TGT CGA GG	
TUA 1F	CTA GTG ACA CAA CAG TAG GTG C	qPCR normalisation gene primer pair
TUA 1R	GAA ACA GTT GCC GAT AAC CAC C	
TUA 2F	GGC CAT TAT ACA GTT GGG AAG G	qPCR normalisation gene primer pair
TUA 2R	ACA ATG AGC CCA AAC CAG AAC C	
TUA 3F	CTA TCC TTC CCC TCA GGT ATC G	qPCR normalisation gene primer pair
TUA 3R	AAG CAC AGC CAC GTC TGT ATG C	
TUA 4F	TTT GTT GAC TGG TGC CCA ACG	qPCR normalisation gene primer pair
TUA 4R	AGA GAA CAC CTC AGC AAC TGC	
ACT 1F	CAG AGT ATA TGG CTT CCT TC	qPCR normalisation gene primer pair
ACT 1R	GGC GAG GAA AAG AAT GAT TTG C	
ACT 3F	GGT GGG ACA TCC TTT TGG ATA G	qPCR normalisation gene primer pair
ACT 3R	CTC GTT ATC GCC TTC TCA C	
ACT 4F	GAC GTT CAA TAC ACC CGC TA	qPCR normalisation gene primer pair
ACT 4R	CAT AGA TAG GGA CGG TAT GG	
ACT 5F	CAT CTA GGA CCA GCT CAT CT	qPCR normalisation gene primer pair
ACT 5R	GAA TCG CTC AGC ACC GAT	
GLO ^T qPCR F	TAG CTC CAG TAA GAT GCA TGA T	qPCR <i>GLO^T</i> expression primer pair
GLO ^T qPCR R	TCT GAG CTC AAT CTG CAT ATT G	
GLO qPCR F	ACT CCG TTG ATT AAC ATC TTG GAT G	qPCR <i>GLO^T</i> expression primer pair
GLO qPCR R	CAA GTG CCT GAG CTC AAT TTG C	
GLO in situ probe F	CAT GAA TAT TGC AGC CCT AAA A	<i>GLO in situ</i> hybridisation probe sequence (Cook 2002)
GLO in situ probe R	CCA TTT TCC ATT TCC CCG CCT	
GLOT in situ probe F	CCT GAA ATA TCA ACC TTG AGC	

GLOT in situ probe R	CTT ATA ATA AAA GGT ACG TCC AGG	<i>GLO^T in situ</i> hybridisation probe sequence
GLO promoter F	AAG GTC TCA GGA GCT TAC TCA TTG ATG TAA CGC ATC	Used in the generation of the promoter-reporter GLO construct
GLO promoter R	AAG GTC TCA CAT tCT CTT TCT TTT TCT CTT CCT CTC	
GLOT promoter F	AAG GTC TCA GGA GTT TGA GCT TTT TCT CAC ACA ATG G	Used in the generation of the promoter-reporter GLO ^T construct
GLOT promoter F	AAG GTC TCA CAT tCT TCC TTC TTA TCT CTC TAG CT	
PI promoter part 1 version 1 F	AAG GTC TCA GGA GGC TCT AGC GGA ACC ACT ACT C	Used in Type IIS cloning
PI promoter part 1 version 1 R	AAG GTC TCG TCT gTT CCG TAC TAT GTT TTC TTC	Used in Type IIS cloning
PI promoter part 2 version 1 F	AAG GTC TCA cAG ACC AGA GGT TAA TTA AAC GAC	Used in Type IIS cloning
PI promoter part 2 R	AAG GTC TCC CCA TCT TTC TCT CTC TAT CTC TTT C	Used in Type IIS cloning
PI promoter part 2 R a	AAG GTC TCA CAT tCT TTC TCT CTC TAT CTC TTT CT	Used in Type IIS cloning
PI promoter part 1 version 2 R	AAG GTC TCT cGT CTC TTC CGT ACT ATG TTT TCT TC	Used in Type IIS cloning
PI promoter part 2 version 2 F	AAG GTC TCA GAC gAG AGG TTA ATT AAA CGA CAC	Used in Type IIS cloning
PI promoter part 2 R (a)	AAG GTC TCA CAT tCT TTC TCT CTC TAT CTC TTT CT	Used in Type IIS cloning
GLO cDNA part 1 F	AAG GTC TCG ATG GGT AGA GGA AAA ATA GAG ATA A	Used in Type IIS cloning
GLO cDNA part 1 R	AAG GTC TCG AcA CCT GAG CAT CAC ATA AAA C	Used in Type IIS cloning
GLO cDNA part 2 F	AAG GTC TCG GTg TCC CTT GTT ATT TTT GCC AAC	Used in Type IIS cloning
GLO cDNA part 2 R	AAG GTC TCT ATT TAA ATC CTC TCC TGT AAA TTT GGC TG	Used in Type IIS cloning
GLOT cDNA part 1 F	AAG GTC TCG ATG GGG AGA GGA AAG GTA GAG	Used in Type IIS cloning
GLOT cDNA part 1 R	AAG GTC TCA cAC CTG AGC ATC ACA CAA AAC CG	Used in Type IIS cloning

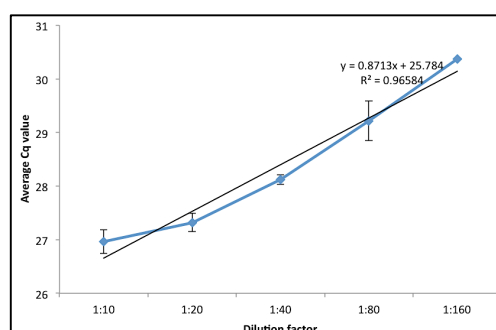
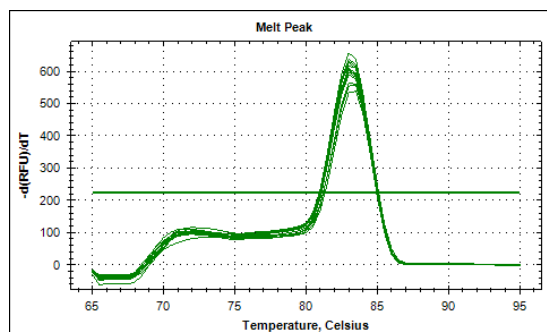
GLOT cDNA part 2 F	AAG GTC TCA GGT gTC CCT TAT TAT TTT CTC TAG CTC	Used in Type IIS cloning
GLOT cDNA part 2 R	AAG GTC TCC TGT ACA CGG TAG GAG CAA GGC	Used in Type IIS cloning
PI cDNA part 1 F	AAG GTC TCG ATG GGT AGA GGA AAG ATC GAG	Used in Type IIS cloning
PI cDNA part 1 R	AAG GTC TCT GaT CTC GGA CTT TGT CGA GGC	Used in Type IIS cloning
PI cDNA part 2 F	AAG GTC TCA GaT CAC CAG ATG GAG ATC CTT ATA	Used in Type IIS cloning
PI cDNA part 2 R	AAG GTC TCA TGA TCA ATC GAT GAC CAA AGA CA	Used in Type IIS cloning
GLO terminator F	AAG GTC TCT AAA TAT AAA CTC AAA ACA AAT TGT CG	Used in Type IIS cloning
GLO terminator R	AAG GTC TCA AGC GAA TGA TTC CGC TGA TCC TAC ATC	Used in Type IIS cloning
GLOT terminator F	AAG GTC TCG TAC AGC CGC TTC AAC CAA ATT TAC	Used in Type IIS cloning
GLOT terminator R	AAG GTC TCA AGC GCT CCG TAA CTA TGA CAT CTT TAA ACG G	Used in Type IIS cloning
PI terminator F	AAG GTC TCG ATC ATC GAG ATT TTA TAA TCT CAT CC	Used in Type IIS cloning
PI terminator R	AAG GTC TCA AGC GGG ATT AAC CAA AAT ATT TCC TCT CC	Used in Type IIS cloning
GLO O.E vector F	AGA GAT GGG TAG AGG AAA AAT AG	Primers used to amplify coding region of <i>P. vulgaris GLO</i> that was used in the overexpression construct
GLO O.E vector F	TTT TGA GTT TAT ATT TAA ATC CTC TCC TG	
PI Y2H F	GGTAGAGGAAAGATCGAGATAAAG	Primers used to amplify coding region of <i>A. thaliana PI</i> that was used in 2YH experiments
PI Y2H R	GATGATCAATCGATGACCAAAGAC	
AP3 Y2H F	GCGAGAGGGAAGATCCA	Primers used to amplify coding region of <i>A. thaliana AP3</i> that was used in 2YH experiments
AP3 Y2H R	TAATTATTCAAGAAGATGGAAGG	

Amplification curves and melt peaks for qPCR primers:

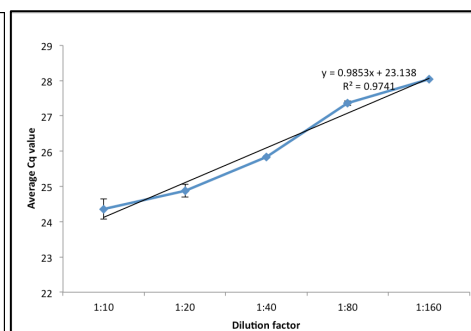
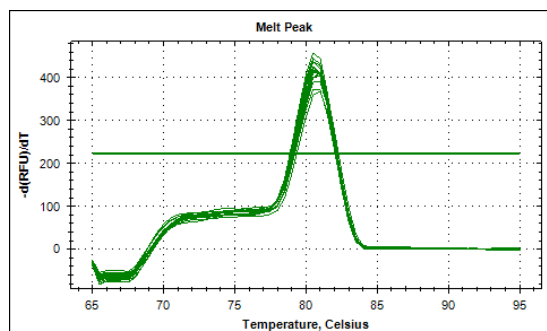
PP2A 1F1R



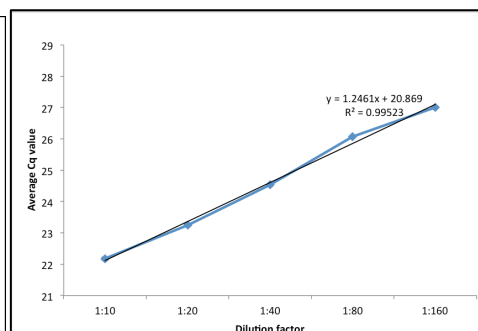
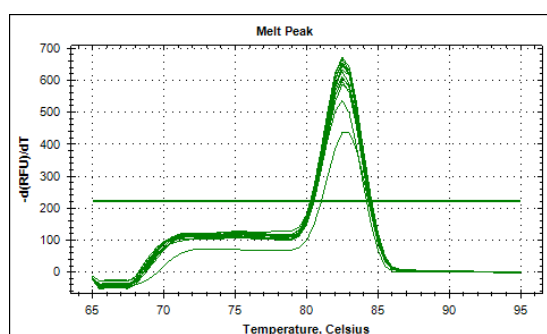
PP2A 2F2R



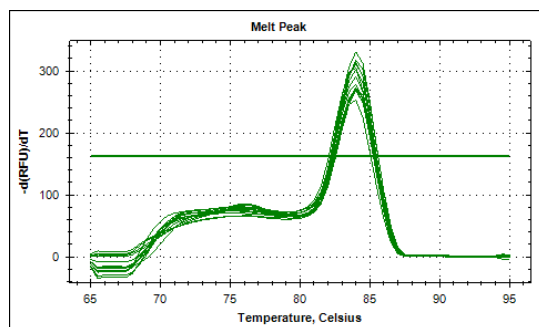
PP2A 3F3R



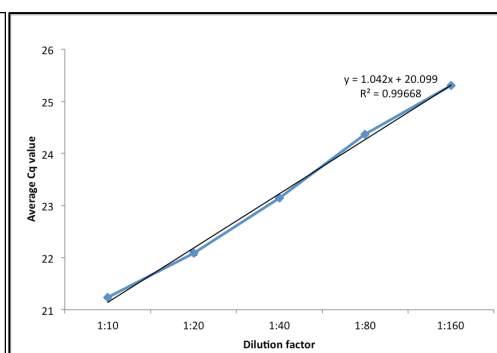
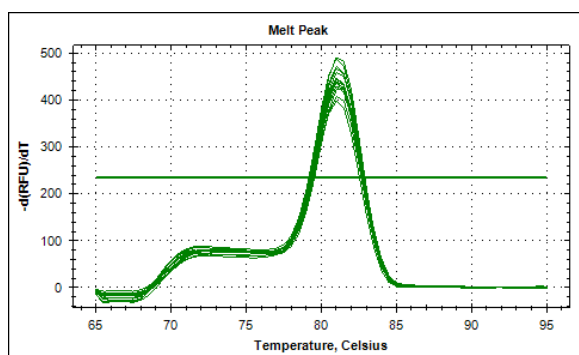
TUA 1F1R



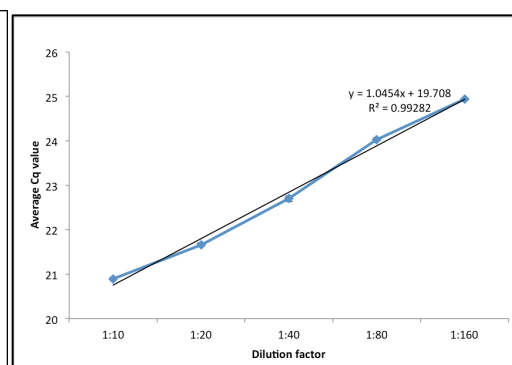
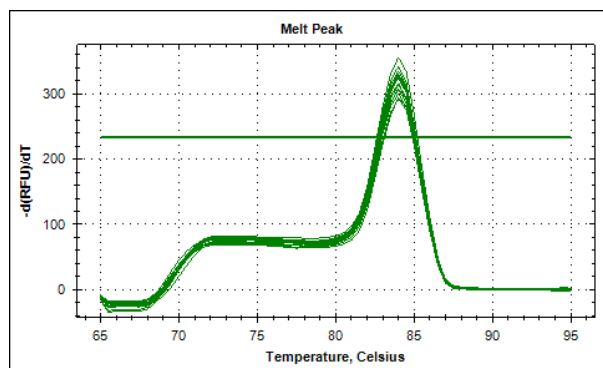
TUA 2F2R



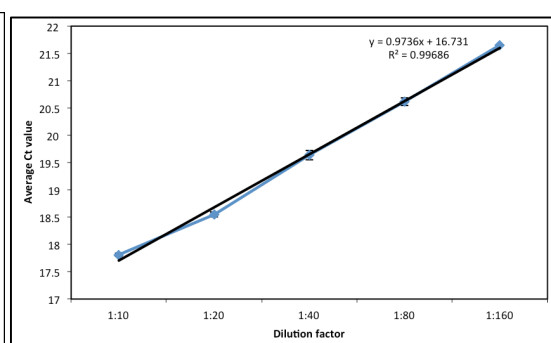
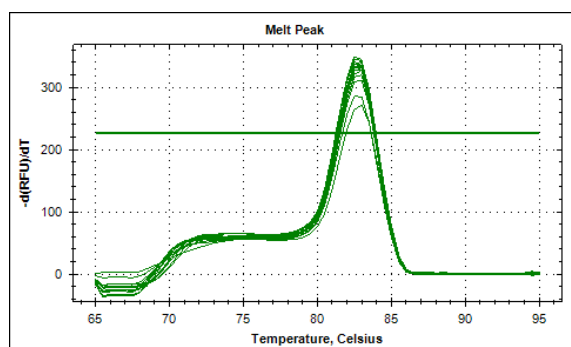
TUA 3F3R



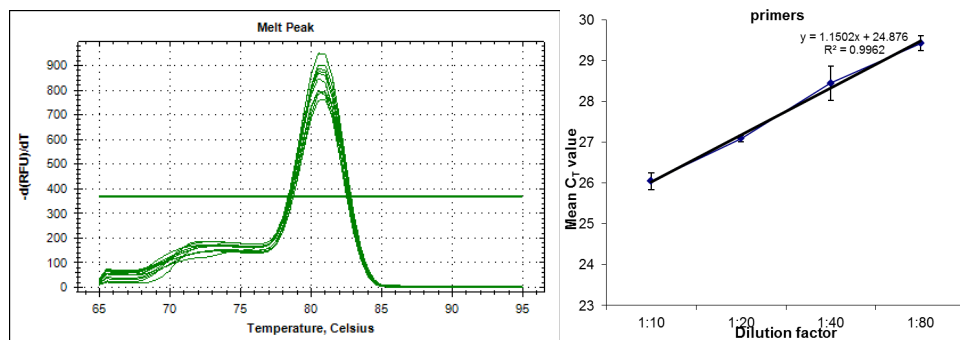
TUA 4F4R



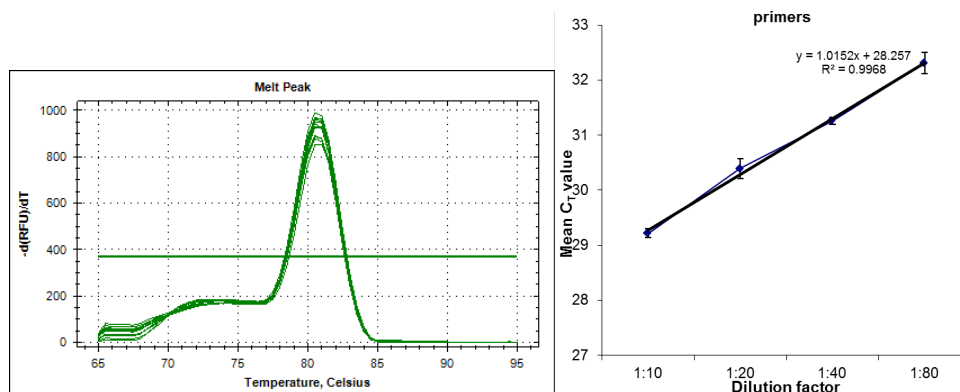
ELF1α 2F2R



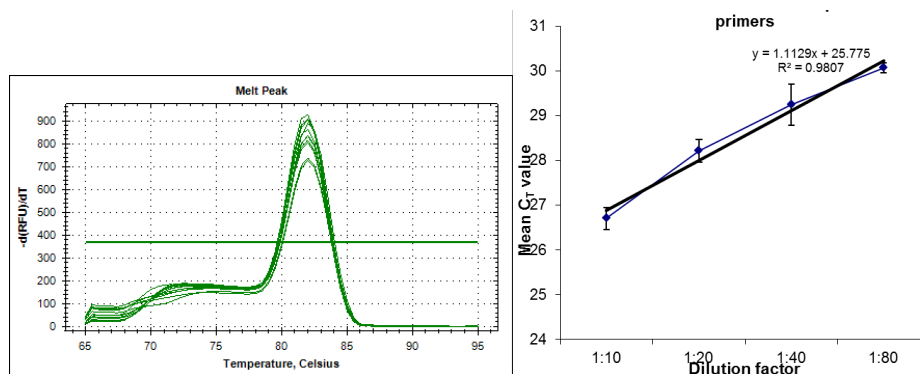
ACT1F1R



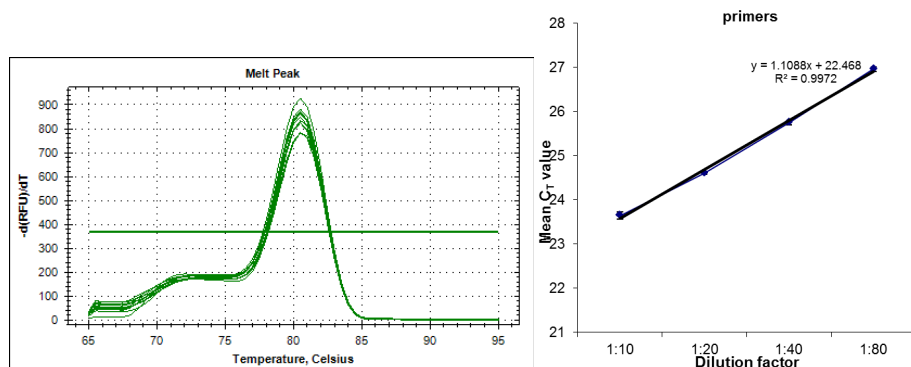
ACT3F3R



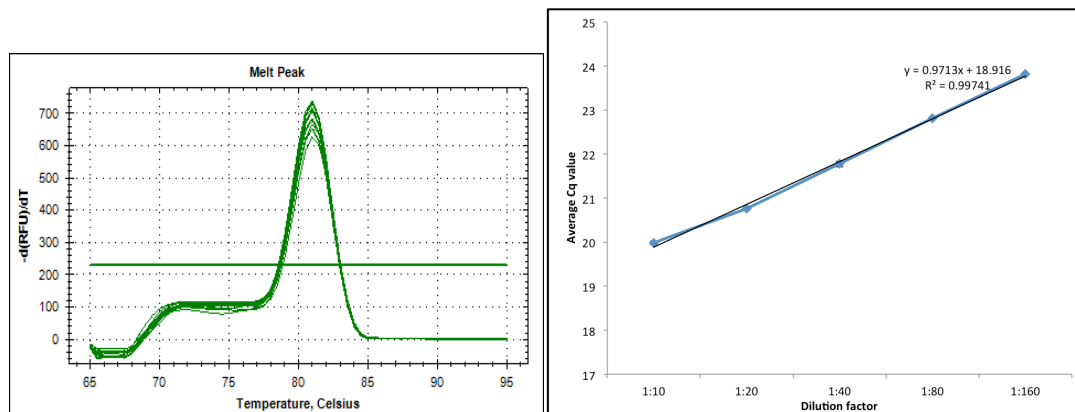
ACT4F4R



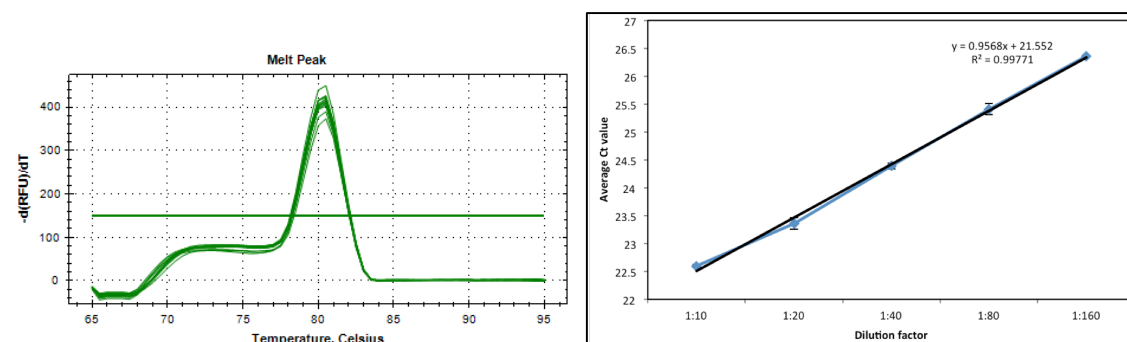
ACT5F5R



GLO



GLO^T



REFERENCES

- Andersen, C. L., Jensen, J. L., & Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research*, 64(15), 5245-50.
- Angenent, G. C., Franken, J., Busscher, M., van Dijken, A., van Went, J. L., Dons, H. J. M., & van Tunen, A. J. (1995). A novel class of MADS-box genes is involved in ovule development in petunia. *Plant Cell*, 7(10), 1569-82.
- Ariño, J., Pérez-Callejón, E., Cunillera, N., Camps, M., Posas, F., & Ferrer, A. (1993). Protein phosphatases in higher plants: multiplicity of type 2A phosphatases in *Arabidopsis thaliana*. *Plant Molecular Biology*, 21(3), 475-85.
- Barrett, S. C. (1978). Heterostyly in a tropical weed: The reproductive biology of the *Turnera ulmifolia* complex (Turneraceae). *Canadian Journal of Botany*, 56(15), 1713-25.
- Bateson, W., & Gregory, R. P. (1905). On the inheritance of heterostylism in *Primula*. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 76(513), 581-586.
- Becker, A. & Theißen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution*, 29(3), 464-89.
- Belles-Boix, E., Hamant, O., Witiak, S. M., Morin, H., Traas, J. & Pautot, V. (2006). *KNAT6*: an *Arabidopsis* homeobox gene involved in meristem activity and organ separation. *Plant Cell*, 18(8), 1900-7.
- Berbel, A., Navarro, C., Ferrandiz, C., Canas, L.A., Beltran, J., & Madueno, F. (2005). Functional conservation of PISTILLATA activity in a pea homolog lacking the PI motif. *Plant Physiology*, 139(1), 174-85.

- Bodmer, W. F. (1958). Natural crossing between homostyle plants of *Primula vulgaris*. *Heredity*, 12, 363-70.
- Bodmer, W. F. (1960). The genetics of homostyly in populations of *Primula vulgaris*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 242(696), 517-549.
- Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes. *Development*, 112(1), 1-20.
- Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E. M. & Smyth, D. R. (1993). Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development*, 119(3), 721-743.
- Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. & Werner, T. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21(13), 2933-2942.
- Causier, B., Schwarz-Sommer, Z. & Davies, B. (2010). Floral organ identity: 20 years of ABCs. *Seminars in Cell and Developmental Biology*, 21(1), 73-79.
- Chen, X., Truksa, M., Shah, S., & Weselake, R. J. (2010). A survey of quantitative real-time polymerase chain reaction internal reference genes for expression studies in *Brassica napus*. *Analytical Biochemistry*, 405(1), 138-140.
- Clough, S. J. & Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal*, 16(6), 735-743.
- Cocker, J., Webster, M. A., Li, J., Wright, J., Kaithakottil, G. G., Swarbrek, D. & Gilmartin, P. M. (2015). *Oakleaf*: an *S* locus-linked mutation of *Primula vulgaris* that affects leaf and flower development. *New Phytologist*, 208(1), 149-161.
- Coen, E. S. & Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature*, 353(6336), 31-37.

- Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H. J. M., Angenent, G. C., & van Tunen, A. J. (1995). The petunia MADS-box gene *FBP11* determines ovule identity. *Plant Cell*, 7(11), 1859–1868.
- Cook, H., E. (2002). *Homeotic genes and their mutations in Primula vulgaris flower development*. PhD Thesis, University of Leeds.
- Crosby, J. L. (1940). High proportions of homostyle plants in populations of *Primula vulgaris*. *Nature*, 145, 672-673.
- Crosby, J. L. (1949). Selection of an unfavourable gene-complex. *Evolution*, 3(3), 212-230.
- Crosby, J. L. (1958). Outcrossing on homostyle primroses. *Heredity*, 13, 127-131.
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., & Scheible, W. R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology*, 139(1), 5-17.
- Darwin, C. (1877). *The different forms of flowers on plants of the same species*. John Murray.
- Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H., & Sommer, H. (1996). Multiple interactions amongst floral homeotic MADS box proteins. *The EMBO Journal*, 15(16), 4330.
- Davies, B., Cartolano, M. & Schwarz-Sommer, Z. (2006). Flower development: the *Antirrhinum* perspective. *Advances in Botanical Research*, 44, 279–315.
- van Dijk W. (1943). La découverte de l'hétérostylie chez *Primula* par Ch. de l'Écluse et P. Reneaulme. *Nedelandisch Kruidkundig Archief*, 53, 81–85.
- Ditta, G., Pinyopich, A., Robles, P., Pelaz, S. & Yanofsky, M. F. (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Current Biology*, 14(21), 1935–1940.

- Dowrick, V. P. J. (1956). Heterostyly and homostyly in *Primula obconica*. *Heredity*, 10, 219-226.
- Edwards, K., Johnstone, C. & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, 19(6), 1349.
- Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D., Patron, N.J. & Marillonnet, S., (2014). A golden gate modular cloning toolbox for plants. *ACS Synthetic Biology*, 3(11), 839-843.
- Ernst, A. (1933). Weitere untersuchungen zur Phänanalyse zum Fertilitätsproblem und zur Genetik heterostyler Primeln. I. *Primula viscosa*. *Archive der Julius Klaus Stiftung für Vererbungs-forschung Sozialanthropologie und Rassenhygiene*, 8, 1–215.
- Ernst, A. (1936a). Erblchkeitsforschungen an calycanthemen Primeln. *Theoretical and Applied Genetics*, 8, 313–24.
- Ernst, A. (1936b). Weitere untersuchungen zur Phänanalyse zum Fertilitätsproblem und zur Genetik heterostyler Primeln. II. *Primula hortensis*. *Archive der Julius Klaus Stiftung für Vererbungs-forschung Sozialanthropologie und Rassenhygiene*, 11, 1–280.
- Ernst, A. (1955). Self-fertility in monomorphic Primulas. *Genetica*, 27(1), 391-448.
- Ferrero, V., Chapela, I., Arroyo, J. & Navarro, L. (2011). Reciprocal style polymorphisms are not easily categorised: the case of heterostyly in *Lithodora* and *Glandora* (Boraginaceae). *Plant Biology*, 13(s1), 7–18.
- Flanagan, C. A., Hu, Y. & Ma, H. (1996). Specific expression of the *AGL1* MADS-box gene suggests regulatory functions in *Arabidopsis* gynoecium and ovule development. *The Plant Journal*, 10(2), 343–353.
- Franklin-Tong, V., E. (ed). (2008). *Self-incompatibility in flowering plants*. Springer.

- Ganders, F. R. (1979). The biology of heterostyly. *New Zealand Journal of Botany*, 17(4), 607-635.
- Ginzinger, D. G. (2002). Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Experimental Hematology*, 30(6), 503-512.
- Goto, K. & Meyerowitz, E., M. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes and Development*, 8(13), 1548-1560.
- Hanson, M. R., & Köhler, R. H. (2001). GFP imaging: methodology and application to investigate cellular compartmentation in plants. *Journal of Experimental Botany*, 52(356), 529-539.
- Hayta, S., Smedley, M., A., Li, J., Harwood, W., A. & Gilmartin, P., M. (2016). Plant regeneration from leaf-derived callus cultures of primrose (*Primula vulgaris*). *HortScience*, 51(5), 558-561.
- Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., & Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology*, 8(2), 1.
- Heuch, I. (1980). Loss of incompatibility types in finite populations of the heterostylous plant *Lythrum salicaria*. *Hereditas*, 92(1), 53-57.
- Honma, T. & Goto, A. (2000). The *Arabidopsis* floral homeotic gene *PISTILLATA* is regulated by discrete cis-elements responsive to induction and maintenance signals. *Development*, 127(10), 2021-2030.
- Hu, R., Fan, C., Li, H., Zhang, Q., & Fu, Y. F. (2009). Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Molecular Biology*, 10(1), 1.

- Huu, C. N., Kappel, C., Keller, B., Sicard, A., Takebayashi, Y., Breuninger, H., Nowak, M. D., Bäurle, I., Himmelbach, A., Burkart, M., Ebbing-Lohaus, T., Sakakibara, H., Altschmied, L., Conti, E., Lenhard, M. (2016). Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. *eLife*, 5, e17956.
- Immink, R. G., Tonaco, I. A., de Folter, S., Shchennikova, A., van Dijk, A. D., Busscher-Lange, J., Borst J. W. & Angenent, G. C. (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biology*, 10(2), R24.
- Jofuku, K. D., den Boer, B. G. W., Van Montagu, M. & Okamoto, J. K. (1994). Control of Arabidopsis flower and seed development by the homeotic gene *APETALA2*. *Plant Cell*, 6(9), 1211–1225.
- Kempin, S. A., Savidge, B., & Yanofsky, M. F. (1995). Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science*, 267(5197), 522.
- Kohn, J. R. & Barrett, S. C. H. (1992) Experimental studies on the functional significance of heterostyly. *Evolution*, 46(1), 43-55.
- Krizek, B. A. & Meyerowitz, E., M. (1996). The *Arabidopsis* homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. *Development*, 122(1), 11-22.
- Lewis, D., & Jones, D. A. (1992). The genetics of heterostyly. In *Evolution and function of heterostyly* (129-150). Springer.
- Li, J., Webster, M., Furuya, M. & Gilmartin, P. M. (2007). Identification and characterization of pin and thrum alleles of two genes that co-segregate with the *Primula S* locus. *The Plant Journal*, 51(1), 18-31.
- Li, J., Webster, M., Dudas, B., Cook, H., Manfield, I., Davies, B., & Gilmartin, P. M. (2008). The *S* locus-linked *Primula* homeotic mutant *sepaloid* shows characteristics of a B-function mutant but does not result from mutation in a B-function gene. *The Plant Journal*, 56(1), 1-12.

- Li, J., Dudas, B., Webster, M. A., Cook, H. E., Davies, B. H. & Gilmartin, P. M. (2010). *Hose in Hose*, an *S* locus-linked mutant of *Primula vulgaris*, is caused by an unstable mutation at the *Globosa* locus. *Proceedings of the National Academy of Sciences*, 107(12), 5664-5668.
- Li, J., Webster, M. A., Smith, M. C. & Gilmartin, P. M. (2011). Floral heteromorphy in *Primula vulgaris*: progress towards isolation and characterization of the *S* locus. *Annals of Botany*, 108(4), 715-726.
- Li, J., Webster, M., Wright, J., Cocker, J. M., Smith, M., Badakshi, F., Heslop-Harrison, P. & Gilmartin, P. M. (2015). Integration of genetic and physical maps of the *Primula vulgaris S* locus and localization by chromosome *in situ* hybridization. *New Phytologist*, 208(1), 137-148.
- Li, J., Cocker, J.M., Wright, J., Webster, M.A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., van Oosterhout, C. & Gilmartin, P.M. (2016). Genetic architecture and evolution of the *S* locus supergene in *Primula vulgaris*. *Nature Plants*, 2, 16188.
- Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K. & Hake, S. (1994). A *knotted1*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *The Plant Cell*, 6(2), 1859-1876.
- Litt, A. & Kramer, E. M. (2010). The ABC model and the diversification of floral organ identity. *Seminars in Cell and Developmental Biology*, 21(1), 129–137.
- Livak, K. J., & T. D. Schmittgen. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*, 25(4) 402-408.
- Lloyd, D. G., & Webb, C. J. (1992). The evolution of heterostyly. In *Evolution and function of heterostyly* (151-178). Springer.
- Maciver, I. (2010). A quick method for A tailing PCR products. <https://promega.wordpress.com/2010/01/19/a-quick-method-for-a-tailing-pcr-products/>

- Mallona, I., Lischewski, S., Weiss, J., Hause, B., & Egea-Cortines, M. (2010). Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. *BMC Plant Biology*, 10(1), 1.
- Manchado-Rojo, M., Delgado-Benarroch, L., Roca, M. J., Weiss, J. & Egea-Cortines, M. (2012). Quantitative levels of *Deficiens* and *Globosa* during late petal development show a complex transcriptional network topology of B function. *Plant Journal*, 72(2), 294-307.
- Mandel, M. A., Gustafson-Brown, C., Savidge, B., & Yanofsky, M. F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature*, 360, 273-277.
- Manfield, I. W., Pavlov, V. K., Li, J., Cook, H. E., Hummel, F. & Gilmartin, P. M. (2005). Molecular characterization of DNA sequences from the *Primula vulgaris* *S*-locus. *Journal of Experimental Botany*, 56(414), 1177-1188.
- Mast, A. R., Kelso, S., & Conti, E. (2006). Are any primroses (*Primula*) primitively monomorphic? *New Phytologist*, 171(3), 605-616.
- McGonigle, B., Bouhidel, K. & Irish, V. F. (1996). Nuclear localization of the *Arabidopsis* APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. *Genes and Development*, 10(14), 1812-1821.
- Melzer, R., Verelst, W. & Thießen G. (2009). The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in ‘floral quartet’-like complexes *in vitro*. *Nucleic Acids Research*, 37(1), 144-157.
- Meyerowitz, E. M., Smyth, D. R. & Bowman, J. L. (1989). Abnormal flowers and pattern formation in floral development. *Development*, 106(2), 209-217.
- Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., & Yanofsky, M. F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*, 405(6783), 200-203.

- Pfaffl, M. W., Tichopad, A., Prgomet, C., & Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnology Letters*, 26(6), 509–515.
- Pinyopich, A., Ditta, G. S., Baumann, E., Wisman, E., & Yanofsky, M. F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*, 424(6944), 85–88.
- Poulter, N. S., Staiger, C. J., Rappoport, J. Z. & Franklin-Tong, V. E. (2010). Actin-binding proteins implicated in formation of the punctate actin foci stimulated by the self-incompatibility response in *Papaver*. *Plant Physiology*, 152(3), 1274–1283.
- Reddy, V. S., Ali, G. S., & Reddy, A. S. (2002). Genes encoding calmodulin-binding proteins in the *Arabidopsis* genome. *Journal of Biological Chemistry*, 277(12), 9840–9852.
- Richards, A. J. (2003). *Primula*. Timber Press.
- Savidge, B., Rounsley, S. D., & Yanofsky, M. F. (1995). Temporal relationship between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity. *Plant Cell*, 7(6), 721–733.
- Riechmann, J. L., Wang, M., & Meyerowitz, E. M. (1996). DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Research*, 24(16), 3134–3141.
- Roh, K. H., Choi, S. B., Kang, H. C., Kim, J. B., Kim, H. U., Lee, K. R., & Kim, S. H. (2014). Isolation and functional characterization of a *PISTILLATA-1* gene promoter from *Brassica napus*. *Journal of the Korean Society for Applied Biological Chemistry*, 57(6), 759–768.
- Sablowski, R. (2015). Control of patterning, growth, and differentiation by floral organ identity genes. *Journal of Experimental Botany*, 66(4), 1065–1073.

- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., & Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science*, 250(4983), 931–936.
- Serikawa, K. A., Martinez-Laborda, A. & Zambryski, P. (1996). Three *knotted1*-like homeobox genes in *Arabidopsis*. *Plant Molecular Biology*, 32(4), 673–683.
- Shivhare, R., & Lata, C. (2016). Selection of suitable reference genes for assessing gene expression in pearl millet under different abiotic stresses and their combinations. *Scientific Reports*, 6, 23036.
- Silver, N., Best, S., Jiang, J., & Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology*, 7(1), 1.
- Smedley, M. A., & Harwood, W. A. (2015). Gateway®-compatible plant transformation vectors. *Agrobacterium Protocols: Volume 1*, 3-16.
- Smith, M. C. (2014). *Assembly and annotation of sequences surrounding the S locus in Primula vulgaris*, PhD thesis, Durham University.
- de Spiegelaere, W., Dern-Wieloch, J., Weigel, R., Schumacher, V., Schorle, H., Nettersheim, D., Bergmann, M., Brehm, R., Kliesch, S., Vandekerckhove, L. & Fink, C. (2015). Reference gene validation for RT-qPCR, a note on different available software packages. *PloS One*, 10(3), e0122515.
- Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M., & Nguyen, M. (2010). A practical approach to RT-qPCR—publishing data that conform to the MIQE guidelines. *Methods*, 50(4), S1-S5.
- Theißen, G. (2001). Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology*, 4(1), 75-85.
- Thompson, J. D., Pailler, T., Strasberg, D. & Manicacci, D. (1996). Tristyly in the endangered Mascarene Island endemic *Hugonia serrate* (Linaceae). *American Journal of Botany*, 83(9), 1160–1167.

- Tröbner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lönig, W., Saedler, H., Sommer, H. & Schwarz-Sommer, Z. (1992). *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. *The EMBO Journal*, 11(13), 4693-4704.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3(7), 1.
- Webb, C. J. & Lloyd, D. G. (1986). The avoidance of interference between the presentation of pollen and stigmas in angiosperms II. Herkogamy. *New Zealand Journal of Botany*, 24(1), 163-178.
- Webster, M.A. (2005). *Floral morphogenesis in Primula: inheritance of mutant phenotypes, heteromorphy and linkage analysis*. PhD Thesis, University of Leeds.
- Webster, M. A. & Gilmartin, P. M. (2003). A comparison of early floral ontogeny in wild-type and floral homeotic mutant phenotypes of *Primula*. *Planta*, 216(6), 903-917.
- Webster, M. A. & Gilmartin, P. M. (2006). Analysis of late stage flower development in *Primula vulgaris* reveals novel differences in cell morphology and temporal aspects of floral heteromorphy. *New Phytologist*, 171(3), 591-603.
- Webster, M. A. & Grant, C. J. (1990). The inheritance of calyx morph variants in *Primula vulgaris* (Huds). *Heredity*, 64(1), 121-24.
- Wuest, S. E., O'Maoileidigh, D. S., Rae, L., Kwasniewska, K., Raganelli, A., Hanczaryk, K., Lohan, A., Loftus, B., Graciet, E. & Wellmer, F. (2012). Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proceedings of the National Academy of Sciences*, 109(33), 13452–13457.
- Yang Y., Fanning L., & Jack T. (2003). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins APETALA3 and PISTILLATA. *The Plant Journal*, 33(1), 47–59.

Zolman, B. K., Monroe-Augustus, M., Silva, I. D., & Bartel, B. (2005). Identification and functional characterization of *Arabidopsis* PEROXIN4 and the interacting protein PEROXIN22. *The Plant Cell*, 17(12), 3422-3435.